Study of the kinetics of self-assembly of viral nanocapsules at high spatiotemporal resolution

Didier Law-Hine

NNT : 2016SACLS072

THESE DE DOCTORAT
DE
L'UNIVERSITE PARIS-SACLAY
PREPAREE À
L'UNIVERSITE PARIS-SUD

LABORATOIRE DE PHYSIQUE DES SOLIDES

ECOLE DOCTORALE N° 564
PHYSIQUE EN ÎLE-DE-FRANCE

Spécialité de doctorat: PHYSIQUE

Par

Didier LAW-HINE

Sonder la cinétique d'auto-assemblage de nanocapsules virales à haute résolution spatio-temporelle

Thèse présentée et soutenue à Orsay, le 2 Février 2016 :

Composition du Jury :

Mme	Brigitte PANSU	Professeur (Université Paris-Sud, Orsay)	Président
Mr	Vladimir LORMAN	Professeur (Université Montpellier II, Montpellier)	Rapporteur
Mr	Guillaume BROTONS	Maître de Conférences (Université du Maine, Le Mans)	Rapporteur
Mr	Yves LANSAC	Maître de Conférences (Université François Rabelais, Tours)	Examinateur
Mr	Wouter ROOS	Professeur (University of Groningen, Groningen)	Examinateur
Mr	Guillaume TRESSET	Chargé de recherche (Université Paris-Sud, Orsay)	Directeur de thèse

Abstract

Viral assembly is an intriguing topic that can be studied using concepts of soft matter physics. Although huge efforts have been made to synthesize hybrid or non-hybrid supramolecular assemblies with viral proteins, the fundamental mechanisms of self-assembly are yet poorly understood. In particular, the kinetic pathways along which the proteins interact with the genome to form highly symmetrical monodisperse architectures are not completely solved.

State-of-the-art techniques such as Time-Resolved Small-Angle X-Ray Scattering (TR-SAXS) enable an *in situ* visualization of the self-assembly kinetics of viral proteins into virus-like particles. In the first part of this thesis, this technique is seen as the method of choice to probe the kinetics of both self-assembly and disassembly of empty capsids built up from the proteins of the Cowpea Chlorotic Mottle Virus (CCMV). Chemical kinetics models coupled with concepts of SAXS theory are devised in order to extract information about the nature of the reaction intermediates, their structure and their typical lifetime. These models are based on a global fitting of TR-SAXS data collected at synchrotron radiation facilities (SOLEIL and ESRF).

The data reveal that proteins self-assemble through the formation of an intermediate close to a half-capsid that is completed with the addition of proteins into an empty spherical capsid (180-mer). This half-capsid size intermediate is big enough with a lifetime sufficiently long to be detected. The disassembly process shows a different kinetic pathway. A cooperative mechanism involving two detected intermediates that are close to a 70-mer and a 30-mer respectively fits well the data.

The encapsulation of ssRNA with CCMV capsid proteins is also examined in this thesis. At neutral pH where the capsid proteins do not spontaneously assemble *in vitro* into empty spherical capsids, electron microscopy images show that there is a non-negligible population of disordered nucleoprotein complexes that coexist with well-formed spherical viruses. Additionally, TR-SAXS kinetic data suggest that the protein-nucleic acid assembly undergoes a structural reorganization in which the capsid proteins make the nucleoprotein complexes more compact as they simultaneously bind to the RNA. Upon acidification, the particles are well-formed viruses as evidenced by electron microscopy images. These findings suggest that the encapsulation of RNA into well-formed viruses is likely a two-step assembly with a binding step and a reorganization step upon acidification.

Acknowledgments

Je tiens à remercier en premier lieu Guillaume Tresset qui a dirigé cette thèse et qui a guidé mes premiers pas dans le domaine de la recherche pendant trois longues années. Tout au long de ces années, il a toujours été disponible pour répondre à toutes les questions que je me posais. Je le remercie vivement pour son optimisme, sa sympathie, sa confiance et son investissement sur l'ensemble des travaux.

J'adresse également mes remerciements aux rapporteurs de ce manuscrit, Vladimir Lorman et Guillaume Brotons, pour l'intérêt qu'ils ont porté à mes travaux. Merci également à Yves Lansac et Wouter Roos pour avoir accepté d'être examinateurs et à Brigitte Pansu pour avoir présidé ce jury.

Je voudrais remercier Doru Constantin qui m'a beaucoup guidé par ses innombrables remarques pertinentes faites lors des multiples discussions que nous avions eu ensemble. Je le suis particulièrement très reconnaissant pour son très très lourd investissement dans cette thèse.

Je remercie Stéphane Bressanelli et Mehdi Zeghal, qui ont participé aux expériences dans les synchrotrons et aux discussions des résultats. Merci à Stéphane pour les bonnes remarques scientifiques dans le domaine de la virologie et les bonnes paroles à mon égard. Merci à Mehdi pour ses justes contributions lors des discussions et répétitions orales pour les conférences et la soutenance de thèse.

Je remercie également les autres personnes qui ont contribué aux projets scientifiques: Sylvain Prevost, Javier Pérez, Anil Sahoo et Prabal Maiti.

Je tiens à remercier les membres du groupe SOBIO, Françoise Livolant pour ses conseils et sa franchise, Amélie Leforestier pour les mises en garde dans le labo de biochimie, Carine Douarche pour d'autres enseignements techniques et histoires humoristiques. Que dire d'Eric Raspaud pour sa curiosité demesurée de tout sur tout. Je salue son expertise et ses talents d'artiste. Je souhaite également remercier Virginie Bailleux pour son incroyable maîtrise de la pipette eppendorf. Je retiens d'elle son sourire et sa joie de vivre. Enfin, je remercie mon collègue de bureau, Jéril Degrouard pour son support et ses blagues légendaires. Je sais juste que lorsqu'il n'est pas dans le bureau, mon efficacité diminue et mon espérance de vie augmente. Je le remercie pour ses encouragements.

Merci à Sandrine Mariot pour être passée dans le bureau plusieurs fois pour m'avoir remonter le moral. Merci pour les galéjades et également pour m'avoir confié Haddock et Bubulle (paix à son âme) pendant les périodes d'été pour m'avoir tenu compagnie.

Je remercie grandement "Sir" Gireeshkumar Balakrishnan Nair. Thanks for the presence during these years and for having showing remarkable aspects of your personality. Mes remerciements vont également à l'autre post-doc du groupe, Nebraska Zambrano, pour son humour pétillant. Merci aux stagiaires pour m'avoir supporté, pour leur contribution et pour m'avoir rendu moins seul: Clément, Narjès, Magali,

Maelenn et Alisier. Merci aux autres doctorants et postdocs du labo avec qui j'ai beaucoup échangé: Sophie, Vincent, Stéphanie, Elise, Santanu, Nicolo, Pabitra ... et ceux qui ne sont plus là depuis un moment: Sounderya, Sumanta, Zenaida et Hector.

Je souhaiterais exprimer ma gratitude aux informaticiens, Ilan et Loan, pour leur aide et pour leur gentillesse. Je remercie aussi les gens de l'instrumentation, Tahoor, Vincent, Sambath, Jérémie, Yannic pour leur bonne humeur et pour m'avoir accueilli au sein de leur groupe et à leur table au Cesfo. Merci à tous ces gens de l'aile Est, étage 1 pour toute cette ambiance conviviale. Merci également à Marta et Katia de l'aile Est, étage 2 pour leur sympathie et pour d'autres conseils.

Merci à l'administration du LPS pour leur efficacité exemplaire, plus particulièrement Mélanie, Sabine et Véronique qui m'ont bien aidé pour les démarches administratives. Je tiens également à remercier Dominique Chandesris pour s'être chargée de trouver une partie du financement de ma thèse.

Je n'oublie pas non plus mes amis, pour certains de grands dalons du 974. Merci à Xavier G. pour la compagnie au Cesfo, Vincent Y.S (et sa femme Shuyao), Bryan A.S, Arnaud N., pour leur contribution sur des programmes informatiques, Julien L., Nasma S., Hussein I., Aurélie L. (bientôt I.), Olivier W.M.K et Morgane T., Matthieu R. du Puy pour leur présence dans des grands évènements VIP. *Merci zot tout' !!*

Enfin, je remercie ceux qui m'accompagnent au jour le jour, papa et maman, frère et soeur, cousins et cousines, tontons et taties et aux deux petites nièces arrivées entre temps. Merci pour les gestes tendres et affectifs à mon égard.

Et maintenant je vous souhaite une bonne lecture de ma thèse ... Vive la République et vive la Science.

“Lorsque tout semble aller contre toi,
souviens-toi que les avions décollent
toujours face au vent”
—Henry Ford

Contents

List of Figures

List of Tables

Chapter 1

General Introduction

Viruses are biological entities that are widely known for their target specificity and self-replication properties. These entities are incapable of independent metabolic activity and replication. Once a specific host-cell is infected, the viruses are building up using the host-cell machinery components for large-scale reproduction.

From a physics point of view, viruses are nanoparticles that possess remarkable thermodynamic properties. The strategies they develop for self-replicating into high-ordered architectures make them fascinating objects to be studied. It is also striking that the viral assembly often result in the formation of objects of a unique size, and that these complete objects can maintain their infectious activity. The key to understand the high degree of such perfection should lie within the comprehension of the fundamental physical mechanisms of self-assembly and genome packaging. Huge efforts have been made to bring answers to these puzzling problems, but are still far from being completely solved yet.

Viruses have long been considered only as infectious agents leading to serious diseases, and therefore they were thought to be systematically hazardous. But their extremely impressive physical properties have been used for positive purposes. In the context of gene therapy for example, viruses are presumed to serve as nanocontainers for gene delivery because of their target specificity. On the other hand, because of the remarkable properties of viral self-assembly, several applications have been found in the field of nanomaterials. Now, viral proteins are being used as nanoblocks for fabricating complex macromolecular assemblies.

Experimental studies on viral self-assembly have essentially been carried out at the thermodynamic equilibrium. The dynamics of self-assembly, however, is still very challenging. From a theoretical point of view, the dynamics of the self-assembly have mostly been studied with particle-based simulations or more recently with molecular

dynamics simulations. Experimentally, it can only be scrutinized by extracting information during the assembly kinetics.

This thesis intends to present new elements in the specific context of viral self-assembly kinetics. It regroups three different studies of kinetics involving viral particles with Time-Resolved Small-Angle X-ray Scattering (TR-SAXS) technique, which rapidly emerged as a technique of choice for the study of rapid kinetics at the nanoscale.

Chapter 2 will introduce shortly the viruses in a general context and several aspects about research on viruses. A section will present a literature review on some theoretical and experimental studies relatively to the specific topic of viral self-assembly. An other section will expose the viral system that I studied during my PhD thesis, the Cowpea Chlorotic Mottle Virus (CCMV). Numerous studies have already been carried out on this system, and the reasons for this will be also explained.

Chapter 3 will expose shortly the theory of X-ray scattering. An entire section will then specifically focus on the TR-SAXS technique. The mathematical aspects on the data analysis and the models developed in this thesis will also be presented.

A first project carried out on the CCMV system will be summarized in Chapter 4. It will concern the disassembly kinetics of CCMV empty capsids. It will be seen how to make use of TR-SAXS technique to extract both temporal and structural information on the disassembly kinetics. The results from this work have been reported and published in 2015.

Chapter 5 gathers the results from the study of the self-assembly kinetics of CCMV empty capsids. This chapter will illustrate the complexity of such a study. It will also highlight the method we developed on the extraction of useful information from the TR-SAXS data. More complex mathematical tools algorithms have been developed especially for this study.

Finally, Chapter 6 will describe the third project carried out during my thesis concerning the assembly kinetics of RNA packaging by CCMV CPs. This project is still ongoing but the main ideas and the first collected data will be presented briefly.

Chapter 2

Viruses, Self-assembly of viruses, CCMV

2.1 Short introduction on viruses

This section intends to present quickly some general aspects of viruses about their structure and classification. The presentation is far from being exhaustive since much can be said on the diversity of viruses. A particular paragraph attempts to describe some of the aspects and overcoming challenges in viruses from the point of view of applications, especially in the research area of nanomaterials.

2.1.1 Viruses of different shapes

Viruses are of different shapes but most viruses can be classified as "rodlike" or "spherical". Examples of viruses that do not belong to these two classes namely include the mature HIV which depicts a conical shape and the bacteriophages which are composed of an (elongated prolate or icosahedral) head and a (long non-contractile) tail (see Fig 2-1). Viruses also differ from an structural point of view. However, one can consider that viruses consist of at least two main components which are the genetic material, which can be DNA or RNA, and a protein shell called capsid. The role of the capsid is to surround and protect the genetic material as well as targeting specific cells thanks to their unique shape and protein sequence.

It is a striking fact that viruses, rodlike or spherical, are extremely regular in size and shape. In the 1950s, Crick and Watson [3] made the hypothesis that all virus capsids are composed of few identical protein subunits packed together and arranged with a high degree of symmetry. Using the example of the rodlike Tobacco Mosaic Virus (TMV) which revealed a helical symmetry around the nucleic acid [4][5], they suggested that the capsid is constructed by repeating the arrangements of the subunits along the helical axis. In the case of a spherical virus, they argued that the limited

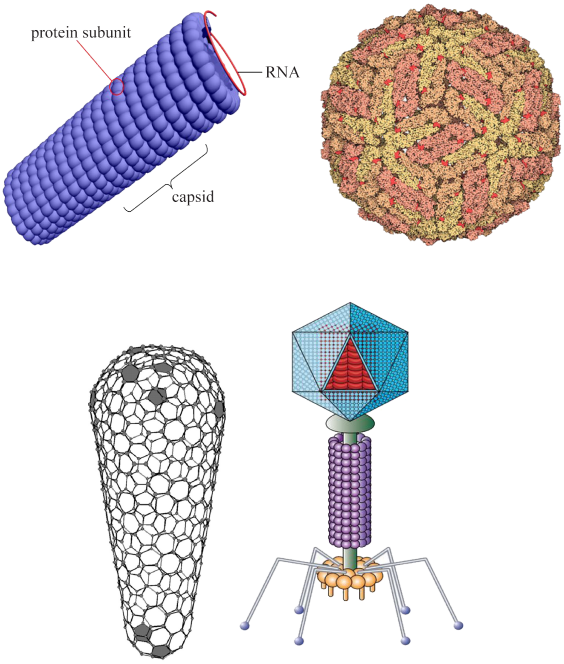


Figure 2-1: Different shapes of virus. Top left: Rodlike virus. Top right: Spherical virus. Bottom left: Cone-like virus (HIV virus for instance). Bottom right : Complex virus (for instance, bacteriophage). Images are adapted from [1] and [2].

volume of the virion capsid should constrain the nucleic acid to code for only a few sorts of proteins. Because of the lack of resolution it has been proposed that these are polyhedron and that for symmetrical reason, asymmetric units must be a multiple of 12.

2.1.2 Icosahedral viruses - Caspar and Klug classification

In the 1960s, Caspar and Klug [6] exhibited more arguments to show that several spherical viruses display an icosahedral symmetry. An icosahedral capsid is composed of 20 facets which are equilateral triangles that comprise at least 3 asymmetric subunits. Given an icosahedron with three subunits per face, 60 subunits can be arranged in a regular manner. However, most virions are composed of more than 60 subunits. To explain how the icosahedral symmetry is maintained in those cases, geometrical arguments were proposed by Caspar and Klug where identical protein subunits interact with slightly different (quasi-equivalent) conformations [6, 7]. The icosahedron can be described as a group of one or two kind of capsomers (pentamers or hexamers) that are groups of protein subunits exhibiting a particular symmetry (5-fold symmetry and/or quasi 6-fold).

Exactly 12 pentamers are required for an icosahedral symmetry. Those pentamers are located at the vertices of the icosahedron. The icosahedral capsid is composed of 60 T subunits (20 facets composed of 3 asymmetric units each), where T is called

the triangulation number which is defined as the square of the distance between two adjacent 5-fold vertices :

$$T = h^2 + hk + k^2 \tag{2.1}$$

where h and k are integer values. Since the total number of protein subunits is $60\,T$ and that there are 12 pentamers, there are $10(T\text{-}1)$ hexamers. Fig 2-2, adapted from [8], illustrates the geometry and construction of icosahedral capsids.

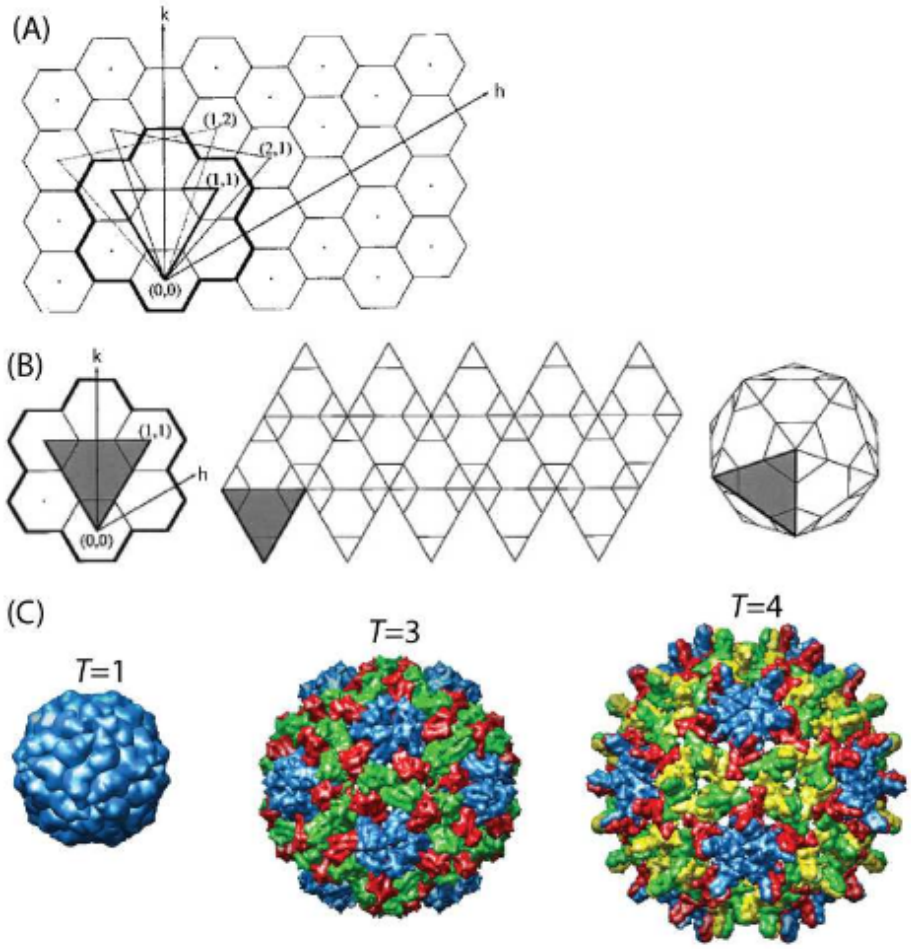


Figure 2-2: Geometry of icosahedral viruses and Caspar Klug theory. **(A)** Construction of different triangular facets depending of $\overrightarrow{h}$ and $\overrightarrow{k}$ lattice vectors. (**B**) Construction of a T=3 lattice. Twenty copies of the triangular facet (left panel) are arranged as shown as in the middle panel. The capsid (right panel) can be obtained by folding the planar drawing in the middle panel. (**C**) Three examples of icosahedral viruses : the T=1 satellite tobacco mosaic virus, the T=3 cowpea chlorotic mottle virus and the T=4 human hepatitis B virus. Structures are shown scaled to actual size. The 60 pentameric proteins are colored in blue. The figure is taken from [8].

Zandi *et al.* [9] proposed a minimal model for equilibrium capsid structure in-

volving protein capsomers with a typical electric potential and minimized internal energy. Using Monte Carlo simulation, they showed that the model reproduced the main structures of icosahedral viruses *in vivo* explaining why some T values are not seen in the Nature.

Some viruses do not follow the Caspar and Klug classification scheme. Some examples of exceptions to the Caspar-Klug rules are the Polyoma Virus [10], the Simian Virus 40 [11] or the dengue virus [12].

2.1.3 Applications in virus research

Viruses are interesting biological objects for applications in many aspects. The fact that they cause many major diseases (AIDS, ebola, ...) is an obvious reason to understand the mechanisms of their propagation in order to stop them. Viruses are still important in life sciences and medecine as they can be used to investigate the cellular machinery (DNA replication, protein transport, immunology,...). They are often used as vectors to introduce genes into the cells. This can be useful to produce these genes in large quantities or to simply study the effect of introducing foreign genes in the cell. Because the viruses can target specific locations within some parts of an host cell, this can be beneficial for gene therapy and drug delivery.

If the use of viruses for diverse applications in the field of nanomedecine is conspicuous, their role in materials science is much less noticeable. Yet, from a molecular point of view, viruses can be seen as highly symmetrical monodisperse nanoparticle which can be used as templates for nanoengineering. The virus can be represented as a molecular container that displays three domains that can be exploited [13](See Fig 2-3) : the interior part, the exterior surface, and the interface between the capsid proteins.

The interior domain of the viral capsid has been studied for encapsulating organic or inorganic particles, making it an "hybrid virus". It can indeed be used as a cargo when substituting the viral genome with the particle to encapsulate. For example, the densely positively-charged interior interface of the Brome Mosaic Virus (BMV) made possible the encapsidation of functionalized gold cores (polyethylene glycol attached to gold nanoparticles, AU-PEG). [14] [15]

The interface between capsid proteins relies on non-covalent interactions that are responsible for the three-dimensional configuration. These are often studied with viruses that exhibit pleomorphism, i.e the ability to alter their shapes and sizes according to the environmental conditions [16]. For example, capsids of different size can be produced through genetic modification [17]. Through chemical modification, it is possible to end up with a range of different structures from icosahedral capsids through tubes to planar sheetlike structures [18].

The exterior surface of the viral capsid is a huge playing field for the design of novel functionalized nanostructures. It can be used to site-specifically attach reactive

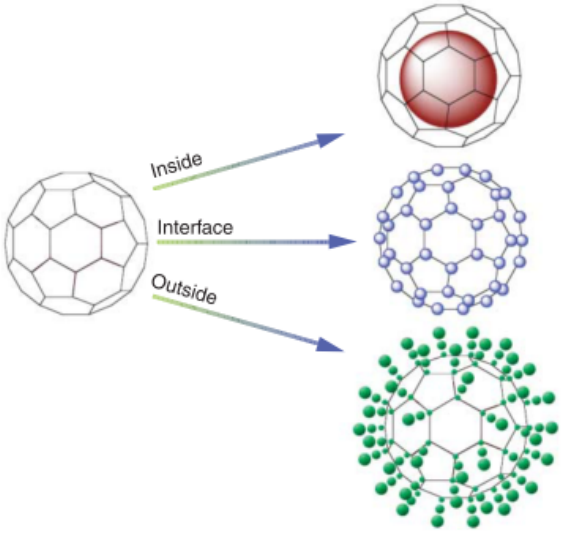


Figure 2-3: Schematic identifying the three accessible interfaces for chemical manipulation in the viral protein cage. The interior part, the exterior surface, and the interface between protein subunits have been used to design new hybrid and functionalized nanoparticles.

groups to the capsids to confer them some chemical properties. As an example, Wang *et al.* used the Cowpea Mosaic Virus to successfully graft fluorophores onto its exterior surface [19], but also azides and alkynes [20]. Multiple attachment of functional groups can be done at the same time. For example, Lee *et al.* [21] used modified M13 virus equipped with peptide groups having affinity for single-walled carbon nanotubes (SWNTs) on one end and peptides capable of nucleating amorphous iron phosphate (a-FePO4) fused to the viral major coat protein. The complex hybrid system was used to fabricate lithium-ion cathodes for batteries.

2.2 Self-assembly of viral capsids : fundamental aspects

A much more fundamental aspect in virus research focuses on the understanding of the mechanisms of the self-assembly of viral capsids. Indeed, although many applications have already been found in the encapsulation of diverse nanoparticles, the fundamental physical mechanisms of self-assembly still remain unsolved. Investigating the physical problem of self-assembly could possibly improve the applications as it could bring more answers to the fundamental questions behind the complexity of interactions taking place during the assembly. Viral assembly usually alludes to the interactions of the capsid proteins and nucleic acids that spontaneously assemble

to form virions. However, if nucleic acids are often required for assembly, in many viruses, the capsid proteins can assemble *in vitro* without them into empty shells depending on conditions such as pH, ionic strength and temperature.

This section is dedicated to a literature survey on the assembly of empty capsids only since Chapter 6 will focus on assembly with RNA. The complexity is such that one may first focus on the interactions between proteins only leading to an empty capsid before investigating the interactions of proteins and nucleic acids leading to a virion. We will consider some experimental works as well as theoretical studies that interrogate the viral capsid assembly problem.

2.2.1 Experimental studies on viral capsid self-assembly

In the literature, two kinds of experiment questionning viral assembly can be distinguished : the self-assembly kinetics and the characterization at equilibrium.

2.2.1.1 Self-assembly kinetics

The kinetics of capsid assembly has been investigated *in vitro* with diverse techniques on different viruses. In 1993, the kinetics of assembly of icosahedral procapsid of phage P22 has been monitored by analyzing the turbidity [22]. Assemblies at different concentrations presented sigmoid curves preceded by a lag time. The estimated rate of assembly was found to be dependent on the concentration and it was suggested that the assembly is determined by a nucleation mechanism. Such mechanism considers monomers or small oligomers in solution that assemble to form a "nucleus". This nucleus would act as the seed in which subunits add rapidly to later form the complete capsid.

The light scattering technique was later used as it provided relatively good time resolution which is an important criterion for studying fast kinetics. This technique also enabled an estimation of the molecular mass. Examples of assembly kinetics followed by light scattering include the study of the Hepatitis B Virus (HBV)[23], the Cowpea Chlorotic Mottle Virus (CCMV)[24], the Human Papillomavirus (HPV)[25], and the Brome Mosaic Virus (BMV)[26]. In the four cases, the assembly profiles also retained a sigmoidal shape that were fitted with a kinetically limiting (KL) model. This model is based on a nucleation-elongation mechanism in which the first step is slow and associated with nucleation, and the subsequent steps are faster steps that correspond to elongation reactions. Elongation reactions suppose that protein subunits add to the nucleus one by one. The association and dissociation rates of the nucleation process were estimated with a fit at the early times of the kinetics which in turn lead to an estimation of the protein association energies.

More recently, the improvement of new techniques using x-rays made possible the study of the kinetics of *in vitro* assembly of empty capsids, namely with the

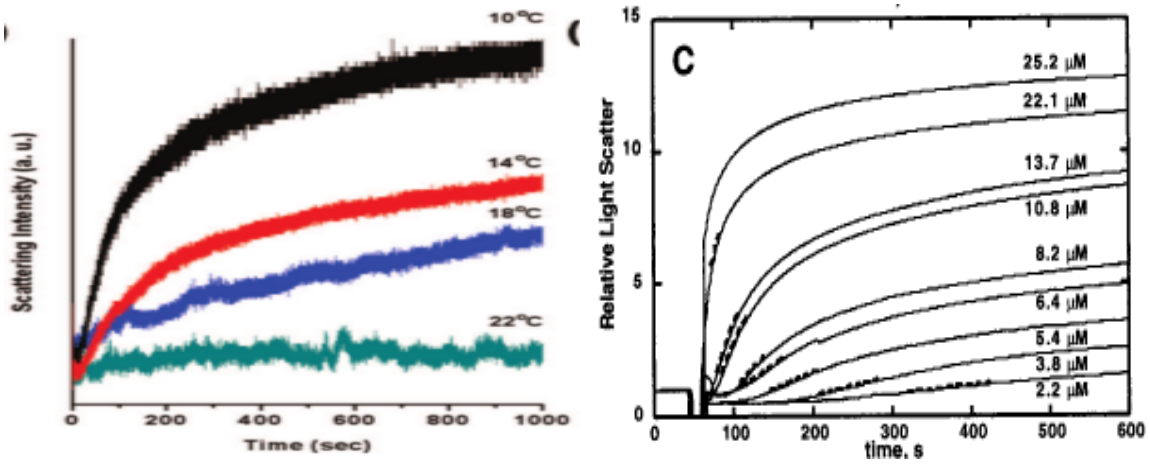


Figure 2-4: Two examples of light scattering experiments on assembly kinetics showing sigmoidal curves. *Left*: Turbidity curves of BMV capsids assembly at various temperatures and protein concentration of 9.4 μM of protein dimers. The capsid assembly rate decreases with temperature. Image from [22]. *Right*: Assembly of varying HBV concentrations by 0.6 M NaCl. Image from [23].

Time-Resolved Small Angle X-Ray Scattering (TR-SAXS) technique. Whereas light scattering only gives temporal information, TR-SAXS enables one to have both spatial and temporal information. This technique of choice has been found suitable to detect and determine the structural enveloppe of potential intermediates during the assembly process of a norovirus [27]. A nucleation-elongation mechanism was first tested and did not fit the data. On the opposite, a cooperative model involving an elongated 11-dimer was found to fit well the data. The TR-SAXS technique together with the global fitting model used in this study will be examined in more details in Chapters 4 and 5, as it is the main technique adopted in this thesis.

2.2.1.2 Characterization at the equilibrium

In the four examples of studies using light scattering technique described in last subsection, the KL model was used to determine the nucleus' size by plotting the concentration dependence of capsid fraction at equilibrium. This method indeed yields the power law coefficient that is often assimilated as the nucleus size. Within the context of the KL model, it was then suggested that HBV capsids (respectively CCMV, HPV and BMV) begin to assemble by forming a trimeric nucleus (respectively pentameric, dimeric and trimeric again).

These experiments however showed the difficulty to characterize assembly pathways because the intermediates are present in very low amounts. Size exclusion chromatography (SEC)[24] and dynamic light scattering (DLS)[26] barely help to distinguish the presence of intermediates. Different strategies are being developped to

address this limitation at the equilibrium. One of them is the use of native mass spectrometry. With this technique, Heck and coworkers [28, 29] identified small oligomers of Hepatitis B virus (HBV) and norovirus under equilibrium with different conditions of pH and ionic strength. This technique is so far the most accurate when dealing with the mass of small oligomers.

2.2.2 Theoretical studies on self-assembly of empty capsids

Theoretical studies on viral assembly are various and therefore this section will introduce some selected concepts related to that problem which have been namely shortly reviewed by Hagan [8] and Roos and coworkers [30]. It was chosen to first deal with some basic thermodynamics principles as those are simply unavoidable when trying to understand the viral assembly. The modeling part of the viral assembly problem will then be discussed, starting with the modeling of the assembly with rate equation models and fitting. Particle-based simulations of the capsid assembly dynamics will then be considered.

2.2.2.1 Thermodynamics of capsid assembly

2.2.2.1.1 Driving forces The 20 amino acids can be catalogued in three groups: charged, polar and hydrophobic amino acids. Capsid proteins are a sequence of amino acids that possess these properties. They have a hydrophobic core, which is not accessible to solvent and a polar surface in contact with the environment. The charges are not distributed along the whole capsid protein but localized at some specific parts which are at the outer surface (usually negatively charged) and the inner surface (usually positively charged).

In a review paper [8], Hagan proposed a short reflexion on the forces driving the assembly of empty capsids. Empty capsid assembly results from protein-protein interactions which are a combination of hydrophobic, electrostatic, van der Waals and hydrogen bonding interactions. Van der Waals interactions and hydrogen bonding occur at short length scales of typically few angstroms. Electrostatic interactions are screened by the presence of salts. The typical scale length in which they operate is the Debye length λ_D. For a monovalent electrolyte,

$$\lambda_D = \sqrt{\frac{\epsilon_0 \epsilon_r k_B T}{2 N_A e^2 \mathrm{I}}} \tag{2.2}$$

where ϵ_0(respectively ϵ_r) are the vacuum (resp. relative) permittivity, k_B the Boltzmann constant, T the temperature, N_A the Avogadro number, e the elementary charge

and I the ionic strength. I is defined as

$$I = \frac{1}{2}\sum_i z_i^2 c_i \tag{2.3}$$

where c_i is the molar concentration of ion i (M, mol/L), z_i is the charge number of that ion. Typically, $\lambda_D \simeq 0.304/\sqrt{I}$ where λ_D is expressed in nanometers (nm) and I is the ionic force expressed in molar (M). With the typical salt concentration used in our study, I(M)=0.5 M, the screening length is $\lambda_D \simeq 0.5$ nm. The hydrophobic interactions occur at similar length scales 0.5 - 1 nm [31][32]. Ceres and Zlotnick [33] showed that the thermodynamics stability of HBV capsids increases with temperature and ionic strength. In their study, the hydrophobic interactions are inferred to be the dominant driving force.

2.2.2.1.2 Law of mass action In the last section " Experimental studies on viral capsid self-assembly", we referred to the work of Zlotnick and coworkers about the characterization at equilibrium of two viruses, HBV [23] and CCMV [34]. They measured the equilibrium concentrations of subunits of different size as a function of the total protein concentration and found two populations of clusters which are the protein subunits and the fully formed capsids. Furthermore, the ratio of the concentrations of free subunits and fully formed capsids seemed to obey quantitatively the law of mass action (LMA). The LMA establishes a relationship between the concentration of an intermediate made of n subunits c_n and that of the subunit itself c_1. At equilibrium, [8, 35]

$$c_n \nu_0 = (c_1 \nu_0)^n e^{-\beta \Delta G_n} \tag{2.4}$$

with c_1 and c_n respectively the concentration of free subunits and intermediates of size n, ν_0 is a reference volume (interpreted as the cell volume used to compute the configuration entropy), $\beta = 1/k_B T$ with k_B the Boltzmann constant, T the temperature and ΔG_n=$G_n - G_1$ the difference in free energy of the intermediate of size n and the free energy of the protein.

Based on the observation that, at the equilibrium, only subunits and capsids are observed, the intermediate concentrations are negligible so that the total concentration can be expressed with the concentration of these two populations

$$c_T = c_1 + N c_N \tag{2.5}$$

With f_c=Nc_N/c_T the fraction of subunits in capsids, one obtain [36][8]

$$\frac{f_c}{1 - f_c} = N(\nu_0 c_T)^{N-1} e^{-\beta \Delta G_N} \tag{2.6}$$

In the limit N $\gg$ 1,

$$\begin{aligned}\frac{f_c^{1/N}}{1-f_c} =& \frac{c_T}{c^*} \\ c^*\nu_0 =& exp\left(\beta\frac{\Delta G_N}{N-1}\right) N^{-1/(N-1)} \approx exp(\beta\Delta G_N/N)\end{aligned} \tag{2.7}$$

with c^* being the pseudo-critical subunit concentration. In the asymptotic limits, Eq 2.7 reduces to

$$\begin{aligned} f_c \approx& \left(\frac{c_T}{c^*}\right)^N \ll 1 \quad , c_T \ll c^* \\ \approx& 1 - \frac{c^*}{c_T} \qquad\quad , c_T \gg c^* \end{aligned} \tag{2.8}$$

Eq.2.8 shows that the fraction of capsid increases with the total subunit concentration c_T and/or the magnitude of binding energy (i.e decreasing c^* since $\Delta G_N < 0$). The transition is also seen to be sharper as the size of the capsid increases.

2.2.2.2 Modeling assembly with rate equation models

2.2.2.2.1 Nucleation-elongation mechanism The empty capsid assembly can be modeled through the use of a system of rate equations that describe the time evolution of concentrations of empty capsid oligomers in solution. Zlotnick and coworkers [23, 37] developed an approach with the simplifications that there is one species of intermediate for each size n, and that only single subunits can bind in each step, leading to the equations of the type

$$\begin{aligned}\frac{dc_1}{dt} =& - f_1 s_1 c_1^2 + b_2 \hat{s}_2 c_2 + \sum_{n=2}^{N} -f_n s_n c_n c_1 + b_n \hat{s}_n c_n \\ \frac{dc_n}{dt} =& f_{n-1} s_{n-1} c_{n-1} c_1 - f_n s_n c_n c_1 \\ & - b_n \hat{s}_n c_n + b_{n+1} \hat{s}_{n+1} c_{n+1}, \quad n = 2...N\end{aligned} \tag{2.9}$$

where c_n is the concentration of intermediates with n subunits , f_n and b_n are respectively the forward (association) and backward (dissociation) rate constants,s_n and $\hat{s}_n$ are coefficients that describe the degeneracy for binding and unbinding [37].

The association and dissociation rate constants are related as

$$b_n = f_{n-1} exp[(G_n^{cap} - G_{n-1}^{cap})/k_B T]/\nu_0,$$

with G_n^{cap} being the interaction free energy of oligomers of size n, and ν_0 the standard state volume. The assembly model therefore requires defining the set intermediate geometries (via the s_n and $\hat{s}_n$ coefficients) and the forward and backward rate constants (or forward rate constants and interaction free energies). This model often refers to

the so-called nucleation-elongation mechanism (seen before in section "Experimental studies on viral capsid self-assembly"), where protein subunits first assemble into a nucleus from which the protein subunits add one by one to form the capsid. This model has been used to simulate and enable comparison with experimental data from the assembly kinetics of HBV [23], the assembly of CCMV into T=3 and pseudo T=2 capsids [34], and the short time kinetics of BMV assembly [26]. It has also been used to fit the assembly of SV40 pentamers on short RNA [38] and calculate the forward rate constants and association energies. In this case, the size of the nucleus was estimated to be between a monomeric (1 pentamer), dimeric (2 pentamers) or trimeric (3 pentamers) nucleus.

2.2.2.2.2 Cooperative mechanism Tresset *et al.* [27] found that a cooperative model, rather than the nucleation-elongation one, better fits the assembly kinetics of norovirus Newbury2 (NB2) virus-like particles. This model looked like the following

$$\alpha \mathrm{D} \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \mathrm{I}$$
$$\beta \mathrm{I} + \gamma \mathrm{D} \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{C}$$

An elongated 11-dimer was found to show up in the assembly kinetics and a cooperative model involving the association of 6 such intermediates with additional dimers to form the capsid was found to fit the data quite well.

2.2.2.3 Particle based simulations

Simulations on capsid assembly dynamics are numerous. This section will discuss only a few works on this field as Hagan [8] reviewed more studies based on simulations of capsids and polyhedral shells assembly. In this section, we will refer to simulations on capsid assembly dynamics that track the positions and orientations of protein subunits.

It is reported that the first dynamical simulations of capsid assembly were carried out by Schwartz and coworkers [39]. They focused on simulations based on the principle of local rules theory [40]. Such theory proposes that the protein subunits take distinct conformations based on the configurations of its immediate neighbors. This theory has been used to explain few puzzling problems about capsid assembly, for example about the Polyomavirus polymorphic capsid assemblies [41], or the difference between T=4 and T=7 geometries with a small perturbation on local rules [42]. The simulations considered a subunit that was modeled as a single sphere with a set of bonds, each of which representing a potential binding interaction with another subunit. They limited their model to the T=1 capsid assembly which corresponded to

the lowest energy configuration. They also used tolerances and flexibility of existing binding interactions that allowed for nonoptimal binding leading to the possibility of having malformed capsids.

Rapaport performed exploratory works by investigating molecular dynamics simulations on polyhedral shells, considering triangular [43] and trapezoidal [44, 45] geometries. Although it was suggested that simple interaction potentials could direct assembly, many nonphysical rules were enforced ensuring that only full capsids would be formed.

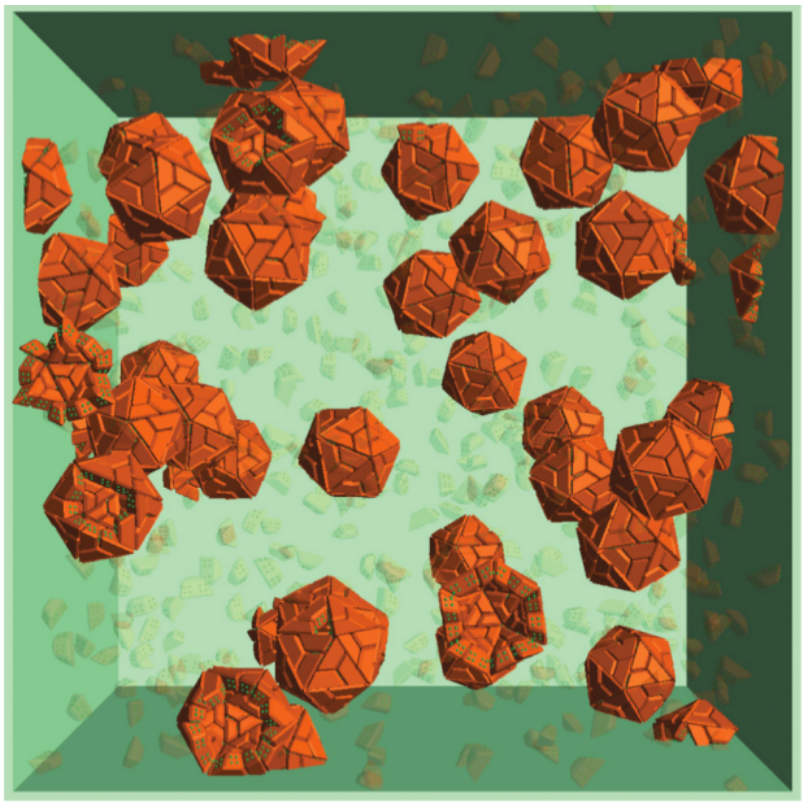


Figure 2-5: Illustration of the simulations performed by Rapaport [45]. The solvent is omitted and particles not in complete shells are shown semitransparently; complete shells that cross periodic boundaries appear open, an artifact of the visualization. Image taken from [45].

Nguyen and coworkers [46] extended these initial ideas on polyhedral shells assembly. They performed molecular dynamics simulation on triangular structures and include hard sphere interactions on each subunit in their model. They found a non-negligible number of non-icosahedral particles accompanying the formation of fully formed capsids. They also predicted a phase diagram depending on subunit concentration and temperature.

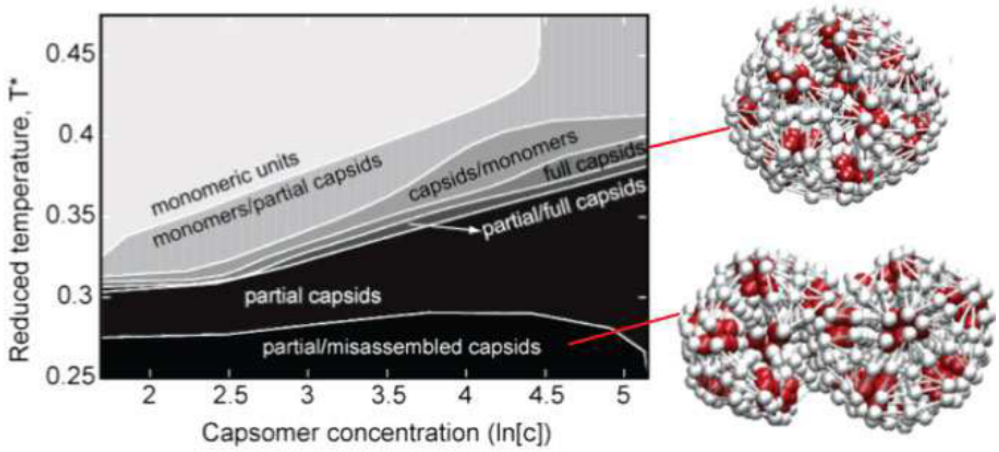


Figure 2-6: Illustration of the simulations performed by Nguyen and coworkers [46]. Assembly products for a 20-subunit as a function of temperature and particle concentration with some representatives structures (well-formed and mis-assembled capsids). Image adapted from [46, 8].

2.3 Cowpea Chlorotic Mottle Virus (CCMV)

Cowpea Chlorotic Mottle Virus (CCMV) is a plant virus that infects the cowpea plant, also known as black-eyed pea. The names "mottle" and "chlorotic" come from the fact that infected leaves develop yellowish spots (see 2-7). This section is dedicated to this particular virus since it is the system that I have studied. The first two subsections focus on the virus itself, while the other subsequent subsections are devoted to the study of the proteins of the CCMV capsid. Indeed, in the first subsections, the CCMV structure will be examined as well as the protocols to show how to isolate and purify the CCMV virions from leaf material. Then, it will be shown how to obtain purified CCMV coat (capsid) proteins by removing the nucleic acid. These proteins exhibit some interesting features that are exploited for nanoengineering purpose.

2.3.1 CCMV structure

CCMV belongs to the bromovirus group of the *Bromovidae* family. The native virus genome is made of four types of single-stranded RNA (ssRNA) molecules: RNA1, 3,171 nucleotides (nt); RNA2, 2,774 nt; RNA3, 2,173 nt; and, RNA4, 824 nt. The first two genomic RNAs are packaged alone, and the third genomic RNA is copackaged with the fourth one, so that each of the three similar CCMV virion contains about 3,000 nt [47, 48].

The X-ray crystallographic structure of the wild-type CCMV has been solved to a 3.2 Å resolution [16] (see Fig 2-8). The information about the protein sequence can be found in a .pdb file under the reference 1CWP.

Figure 2-7: Leaves infected by CCMV displaying yellow spots about 15 days after infection.

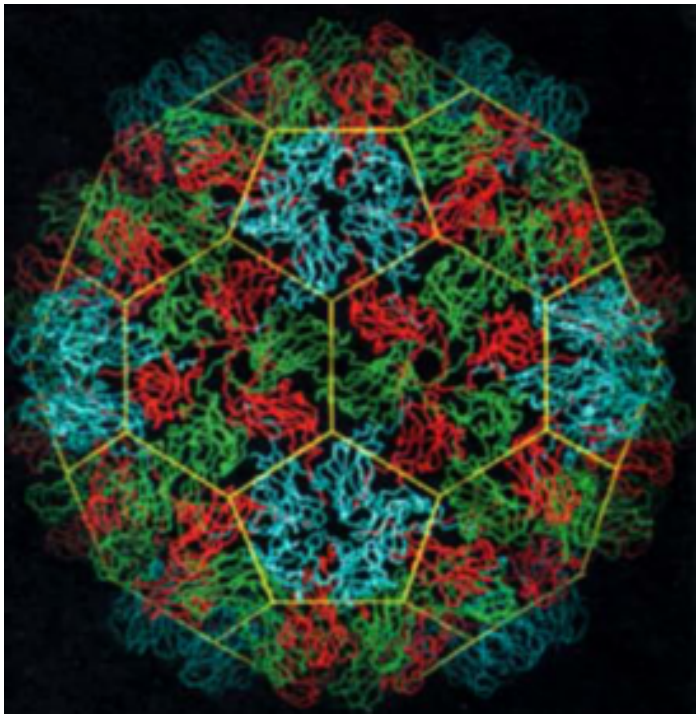


Figure 2-8: Structure of the CCMV capsid. The 60 pentameric proteins are colored in blue (proteins A). The proteins B and C form the hexameric subunits and are colored respectively in red and green. The image is adapted from Speir et al [16].

The capsid is composed of 180 chemically identical coat proteins. The wild-type CCMV (wtCCMV) forms an icosahedral shell with a T=3 quasi-symmetry. Its mass is 180×20 kDa=3.6 MDa. The outer diameter is 280 Å and the protein shell thickness is 50 Å. The proteins are arranged in capsomers consisting of 12 pentamers and 20 hexamers. The morphology of the virion is that of a truncated icosahedron with three chemically identical subunits (monomers), conventionally denoted as A, B, and C. The monomers A form the pentamers of the capsid and the monomers B and C, the hexamers. Every protein subunit is composed of 190 amino acids. The amino acids 1-25 form the N terminal which is located on the inside of the capsid [49]. Since these residues are not ordered in the icosahedron formed by the capsid, they are not seen in the crystallographic structure. The N terminal plays an important role in the interaction with the RNA [50, 51] since the RNA is negatively charged and the N terminal possesses 10 cationic residues.

In solution, when the virus is completely disassembled, the subunits exist as dimers [52, 53]. The fact that dimers and not monomers have been found experimentally in solution has been interpreted with structural considerations that involve the strong interactions between a subunit A and a subunit B on the one hand, and two subunits C on the other hand [16]. As a result, the formation of a capsid can be seen as the assembly of 60 dimers A-B and 30 dimers C-C that exist in solution.

2.3.2 CCMV production and isolation

In 1967, Bancroft *et al.* [18] described the first experiments to isolate the CCMV from harvested leaves. Since then, efforts have been done to improve the purification of viruses according to the purpose or the use of them. For example, Speir *et al.* [16] reported a purification to provide crystalline material suitable for X-ray diffraction measurements. Zhao *et al.* [53] developped a purification protocol using a molecular biology approach. Indeed, in their protocol, CCMV virions were not obtained from infected leaves but from bacteria (*Escherichia Coli*) via *in vitro* assembly using protein expression and RNA transcription. This approach allowed them to construct different mutants of the virus.

2.3.2.1 Production and purification of CCMV

This paragraph describes the protocol used in our laboratory to grow and isolate CCMV virions from harvested infected leaves. It is adapted from that of Ali and Roossinck [54] and Aragones *et al.* [55].

Planting of the cowpea plants

The virus was grown in cowpea plants (*Vigna unguiculata*).
1) Pots (40 to 80) were filled with soil and the cowpea seeds (2-3 seeds per plot) were

planted 1-2 cm deep in the soil. Water was not added directly on the pots, but rather on the trays for the first two days.
2) The plants were grown for about 10 days in a growth room with two lamps hanging 1 meter above the plants. The illumination was set to 12 h per day. Water was added every day on the pots afterwards.

Plant infection

The growth of plant consists in the apparition of primary leaves and then secondary leaves. When the primary leaves were present (about 10 days after planting), they were dusted with carborundum to abrase the leaves to inoculate the virus.
1) The inoculation solution was prepared from a mixture of previously infected cowpea leaves. This mixture was prepared using mortle and pestle. Infected leaves (5-6 leaves) were mixed with 15 mL of distilled water.
2) The primary leaves were inoculated by smoothly rubbing them with carborundum and the inoculation solution. Not too much pressure on the leaves was applied to avoid leaf damage. About 50 μL of solution was applied in one leaf.
3) Immediately after infection, the pots and trays were abundantly filled with water. Note: After about a week, secondary leaves appear with some yellow spots on some of them.
4) 10 to 14 days after inoculation, the plants were harvested by cutting only the leaves. All the leaves (primary and secondary) were cut even if some of them did not show yellowish spots. The infected leaves were then stored at -80° C.

CCMV isolation procedure

The buffers (see *Equipment and buffers* paragraph for more details) were autoclaved before use at 120 °C for 30 min and all steps were carried out at 4 °C or in ice.
1) 70 g of infected cowpea leaves were blended with 150 mL of buffer A (0.15 M sodium acetate, pH 4.8). A nylon filter membrane was used to help mixing by removing larger plant debris. The blending can last up to 20 min but is an important step for facilitating the homogenization.
2) 150mL of ice-cold chloroform was added and the macerate was stirred for 1 h followed by a 10 min centrifugation at 10,000×g.
3) CCMV virions in the aqueous upper phase were precipitated by the addition of NaCl to a final concentration of 20 mM and 8% polyethylene glycol (MW 8000).
4) After stirring for 1 h, the suspension was again centrifuged at 10,000×g for 10 min.
5) The pellet was resuspended in buffer B (0.05 M sodium acetate, pH 4.8) by stirring for 1 h.

6) After centrifugation at 8,000×g for 10 min, the pellet was discarded and the supernatant was centrifuged through a 20% (w/v) sucrose cushion at 150,000×g for 2 h.
7) Immediately after the centrifugation, the supernatant was discarded and the virions in the pellet were resuspended in buffer B and stored at -80° C. The resuspension process could take several hours. One can even stir the suspension in buffer B overnight in cold room in sterile environment. The final aliquot was stored in several tubes (1 mL).
8) The absorbance ratio A_{260}/A_{280} was systematically measured to check both the purity and the concentration of the virions (see section *Characterization*).

From 70 g of infected leaf material, typically 5-10 mg of CCMV virions are obtained at the end of the purification. The purity is always checked to be A_{260}/A_{280}=1.5-1.7 (see section *Characterization*).

Equipment and buffers

The buffers used in this purification were sterilized before use.
Buffer A : 0.15 M sodium acetate, pH 4.8
Buffer B : 0.05 M sodium acetate, pH 4.8
Two centrifuges were used:
- Sorvall RC 5C Plus centrifuge with two kind of rotors : Sorvall GSA rotor for centrifugation at 10,000×g and Sorvall SS34 rotor for centrifugation at 8,000×g.
- Ultracentrifuge : Beckman LE-70 ultracentrifuge with SW 41 rotor (Swinging bucket rotor) for centrifugation at 150,000×g.
- Ultracentrifugation tubes : Beckman Ultra-clear tubes (14×89 mm) purchased from Beckman Coulter.
- Nanodrop 2000 for UV/Vis spectroscopy

2.3.2.2 Characterization

Two main techniques were used to characterize the purification which are the UV/Vis spectroscopy and the electron microscopy (TEM).

UV/Vis spectroscopy
The UV/Vis spectroscopy is a classic technique to obtain information about the purity and the concentration of viruses. It is based on the absorbance principle. Using the Beer Lambert law it is possible to relate the amount of light absorbed to the concentration of the absorbing molecule. The UV/Vis spectrum of CCMV is typically that of a virus in solution with a maximum absorbance at λ=260 nm and a minimum at λ=240 nm (cf Fig 2-9). The strong absorbance at 230 nm is due to the presence of contaminants which can be EDTA and carbohydrates. When looking

at purity of viruses that is composed of proteins and nucleic acids, one has to focus on the ratio of absorbances A_{260}/A_{280}. The wavelength λ=260 nm corresponds to the wavelength of maximum absorbance of the nucleic acids (RNA and DNA). λ=280 nm is the characteristic wavelength of strong absorbance by proteins. Although CCMV contains 25% of RNA and 75% of proteins in terms of mass, the contribution of the nucleic acid is much stronger than that of the proteins. This explains the maximum near λ=260 nm.

In the case of CCMV, the purification is considered as successful when the ratio A_{260}/A_{280} is 1.5-1.7 [56]. Using the molecular extinction coefficient of CCMV (A=5.8 at λ=260 nm for a concentration of 1 mg.mL^{-1} and optical length l=1 cm), one can calculate the concentration c of the viruses in solution with the relation $c = A_{260}/5.8$.

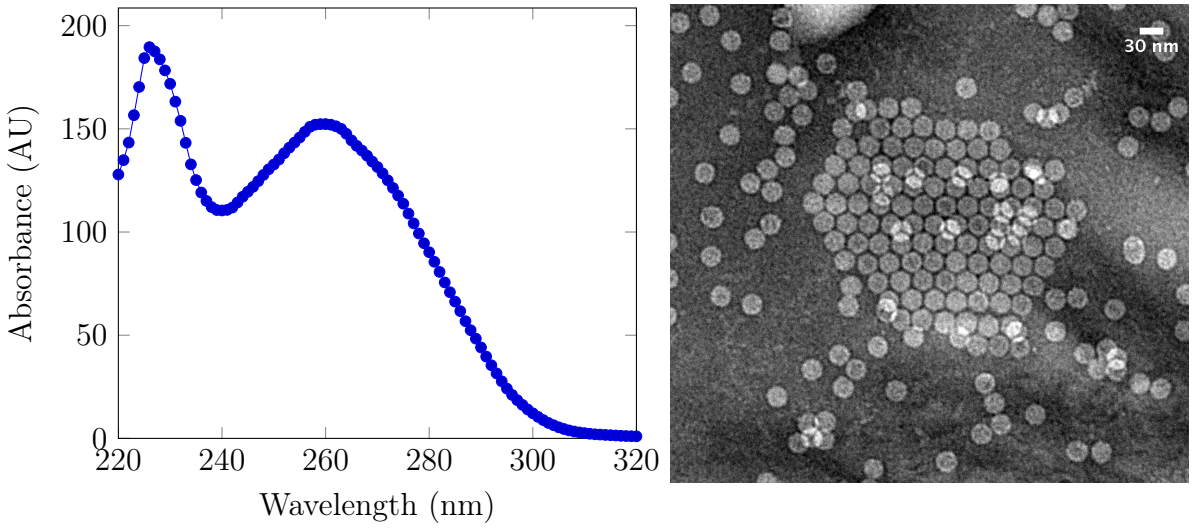


Figure 2-9: Left: Absorbance spectrum of CCMV virions. Right: TEM images of CCMV virions.

Electron microscopy

Electron microscopy allows to visualize directly the particles in solution. It is a useful technique provided that the particles are not too small (typically smaller than 5 nm). In our case, it allowed us to verify if the proteins are assembled or not (formation of viruses/capsids). The protocol is described as follows. Transmission electron microscopy (TEM) was carried out with a JEOL JEM-2010 electron microscope at 200 kV equipped with a Gatan Ultrascan 1000 CCD camera (Gatan, France). 5 μL of sample solution at a protein concentration of $\sim$ 0.1 mg/mL was deposited on a carbon grid treated by glow discharge for 10 min and washed with the edge of filter paper. A drop of 1% (v/w) ammonium molybdate in deioinized water was then incubated on the grid for 30 s. Excess stain was removed as before and the grid was allowed to air dry for at least 10 min.

2.3.3 CCMV coat proteins isolation

Once the virus is purified, the nucleic acids can be removed from the virus and the coat (capsid) protein (CP) subunits can be isolated. Under some conditions, these isolated proteins can interact to assemble and form structures that are similar or close to the spherical capsid. This very special feature has been seen as an interest for research purpose on the assembly of proteins.
In 1968, Bancroft *et al.* [57] first described the re-assembly of the CCMV capsid proteins once the RNA was removed. Protein purification is a critical step for one to consider that the RNA takes no part on the assembly process.

2.3.3.1 Purification of CCMV coat proteins

This paragraph describes the protocol used in the laboratory to purify the CCMV coat proteins and isolate them from the RNA. It is adapted from that of Lavelle and Gelbart [58, 59].

CCMV CP purification

CCMV capsid proteins were purified from CCMV virions obtained from above (section *CCMV isolation*). The equipment and buffers (see *Equipment and buffers* paragraph below for more details) were sterilized before use at 120 °C for 30 min and all steps were carried out at 4 °C.
1) CCMV virions (typically 5 mL of solution) were first dialyzed overnight (17 h) in a 500 mL disassembly buffer (500 mM $CaCl_2$, 50 mM Tris-HCl, 1 mM EDTA, pH 7.5). 1 mM DTT and 0.5 mM PMSF were added in the dialysis bath. CCMV virions were dialysed again for 8h with another 500 mL of disassembly buffer. The buffer is stored for a measurement of absorbance in step 3) as it serves as the blank in the measurement.
2) The solution was then centrifuged for 18 h at 150,000×g. Two 14 mL centrifugal tubes were used : one for the solution and one for the balance.
3) Immediately after the ultracentrifugation, the supernatant was sampled in 1 mL tubes (typically 10-12 tubes). We retain only those that are not contaminated with RNA, i.e, typically those with absorbance ratios $A_{280}/A_{260} \geq 1.45$ (see *Characterization* paragraph below for more details). The second phase at the bottom is yellowish and contains the precipitated RNA. This phase is discarded in the next.
Note: It is important to do step 3) immediately after the end of the ultracentrifugation since the RNA can quickly rediffuse in the solution.
4) 4-5 centrifugation cycles at 4,000×g for 45 min each were performed on these tubes to allow a buffer exchange. The buffer can be either a buffer A (50 mM Tris-HCl, 0.5 M NaCl, 1 mM EDTA, pH 7.5) or a buffer B (50 mM sodium acetate, 0.5 M NaCl, 1 mM EDTA, pH 4.8) to obtain either stable dimers or capsids.

Note: If the CCMV CP dimers are not used for forthcoming experiments, it is preferred to store them as capsids to limit degradation.
5) A systematic measurement of absorbance was done on the final solution to check the concentration and purity of proteins.

With this protocol, typically 3-4 mg of CCMV coat proteins is obtained from 8 mg of virus. The ratio A_{280}/A_{260} is checked to be greater than 1.4 (see section *Characterization*).

Equipment and buffers

Preparation of buffers:
- Disassembly buffer : 500 mM $CaCl_2$, 50 mM Tris-HCl, 1 mM EDTA, pH 7.5. After sterilization, 1 mM DTT and 0.5 mM PMSF was added in the buffer just before the dialysis steps. $CaCl_2$ helps precipitating the RNA, the Ca^{2+} ions combining with the phosphate groups of RNA. EDTA deactivates metal-dependent enzymes in order to prevent damage to proteins. DTT is used to avoid the disulfide bonds of proteins. PMSF is a serine protease inhibitor.
- Buffer A : 50 mM Tris-HCl, 0.5 M NaCl, 1 mM EDTA, pH 7.5. This is used to have stable dimers in solution.
- Buffer B : 50 mM Sodium Acetate, 0.5 M NaCl, 1 mM EDTA, pH 4.8. This is used to have stable capsids in solution.

Equipment:
- Dialysis membrane : Float A Lyzer, 3.5-5 kDa, 5 mL purchased from Sigma-Aldrich.
- Ultracentrifuge : Beckman LE-70 ultracentrifuge with SW 41 rotor (Swinging bucket rotor) for centrifugation at 150,000×g.
- Ultracentrifugation tubes : Beckman Ultra-clear tubes (14×89 mm) purchased from Beckman Coulter.
- Centrifuge : Eppendorf Centrifuge 5804 R with swing bucket rotor A-4-44 for centrifugation at 4,000×g.
- Centrifugation tubes : Amicon Ultra centrifugal filters, 3 kDa, 4 mL purchased from Millipore.
- Nanodrop 2000 for UV/Vis spectroscopy

2.3.3.2 Characterization

To assess the purity of protein, the absorbance at λ=260 nm and λ=280 nm is measured. The ratio A_{280}/A_{260} is a quantitative information about nucleic acid contamination. Verduin [60] calculated that a ratio $A_{280}/A_{260} \geq 1.3$ indicates less than 2.3% (w/w) residual RNA. The absorbance peak at 230 nm is due to the presence of EDTA (cf. Fig 2-10). TEM experiments can be done to ensure the presence of assembled capsids (cf. Fig 2-10).

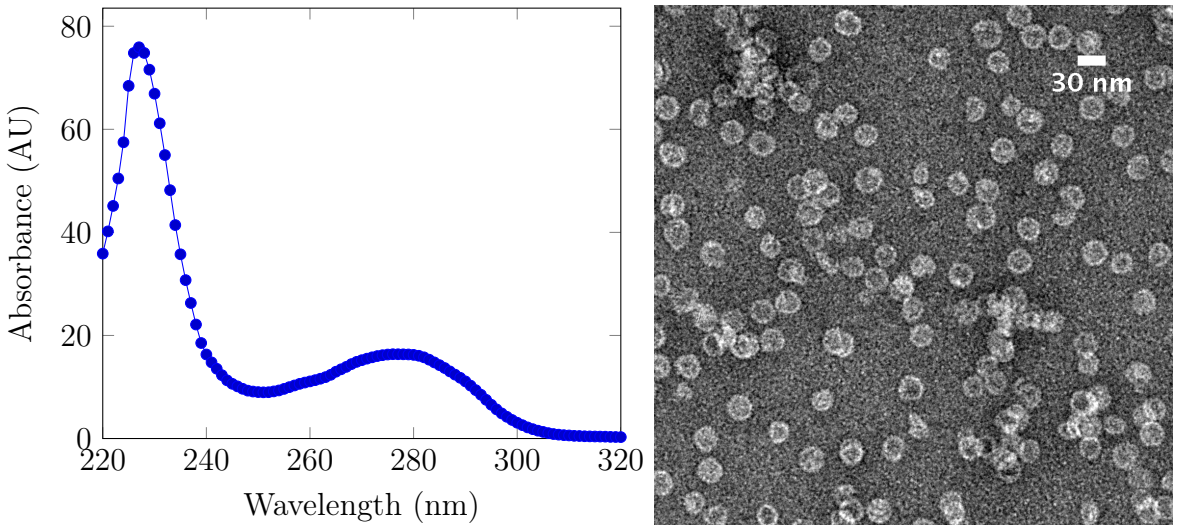


Figure 2-10: Left: Absorbance spectrum of CCMV CPs. Right: TEM images of CCMV CPs.

2.3.4 CCMV coat proteins phase diagram

2.3.4.1 Variety of structures : single-wall and multi-wall shells, tubes, dimers

As seen in last section, the isolated coat proteins can reassemble under some conditions (0.5 M NaCl, pH 4.8) to form spherical capsids. In other conditions of pH and salinity, the capsid proteins can assemble and disassemble reversibly into many other different structures. In 1968, Bancroft *et al.* [57] were first to watchfully examine the self-assembly of CCMV CPs at different pH and ionic strength. Adolph and Butler [61], in 1974, also brought contribution to the equilibrium phase diagram of the CCMV CPs, later revised by Lavelle *et al.* [59] (cf Fig 2-11).

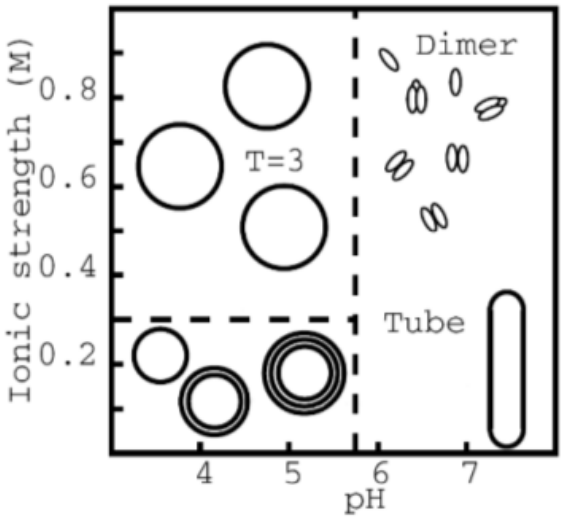
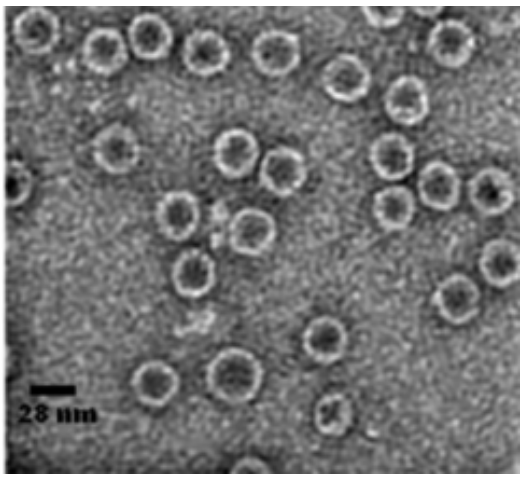


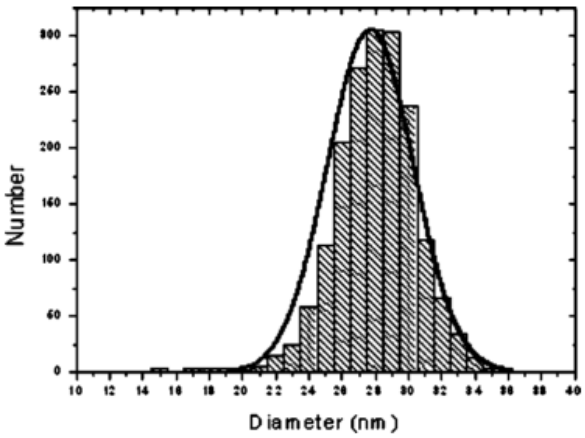
Figure 2-11: Equilibrium phase diagram of the CCMV capsid proteins, adapted from [62, 59]. Protein concentration varies between 0.5 and 1.0 g/L. Temperature is 5°C

The most common structures are the CCMV capsids (single-walled shells T=3) and the dimers. The dimers cannot be seen with electron microscopy because of their small size. T=3 capsids and dimers are namely characterized by multi-angle light scattering and size exclusion chromatography techniques [24] which yield their molecular mass (3.6 MDa for T=3 capsids and 40 kDa for dimers). T=3 capsids can however be characterized with TEM (cf Fig 2-12).

It is only when the ionic strength is very low, that a large number of polymorphs are observed (cf Fig 2-13). At low pH and low salinity, Lavelle et al.[59] observed multi-wall shells which are most frequently composed of two and three walls. There are rarely some four-wall and five-wall shells. The two and three-wall shells are generally well-formed, while the shells with more than three walls are often deformed and incomplete. At very low ionic strength, tubes are occasionally found in coexistence with shells (see Fig 2-13). Their diameter is irregular but they tend to become longer with increasing pH.

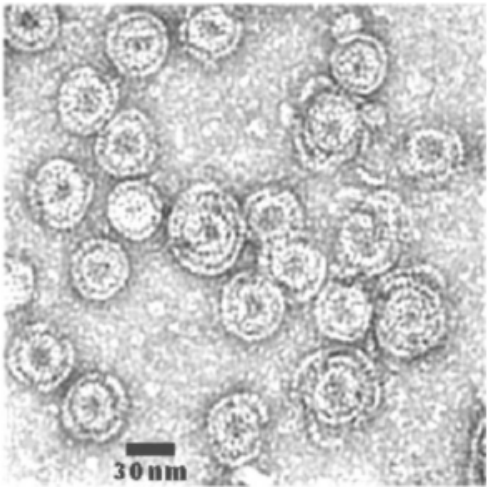


(a) TEM image of single-wall shells in 0.1 M sodium acetate buffer, pH=4.7

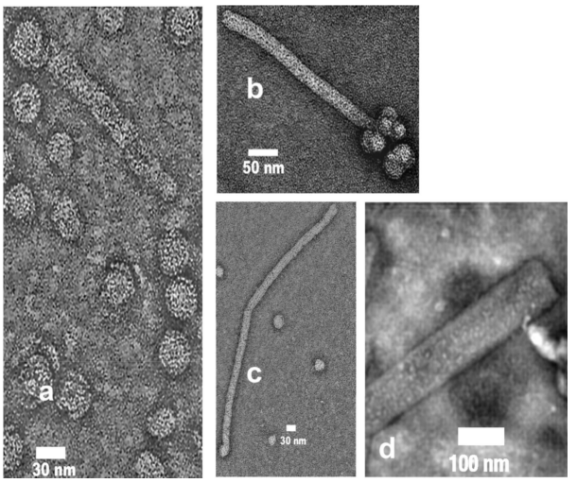


(b) Diameter distribution of single-wall shells. The line is a Gaussian fit to the data, with a maximum at 27.9 nm

Figure 2-12: Single-wall shells and their diameter distribution. The figures are taken from [59]. The protein concentrations were 0.1-0.5 mg/ml, and the assembly was carried out at 4°C.



(a) TEM images of multi-wall shells in sodium acetate buffer. I=0.01 M, pH=4.8



(b) TEM images of tubes obtained with sodium cacodylate buffer and no added salt: (a) pH=5.65, I=0.01 M; (b) pH=6.0, I=0.01 M; (c) pH=7.1, I=0.01 M buffer; (d) pH=7.5, I=0.001 M.

Figure 2-13: Multi-wall shells and tubes observed at low ionic strength. The figures are adapted from [59]. The protein concentrations were 0.1-0.5 mg/ml and the temperature was 4°C.

2.3.4.2 More structures at extreme conditions

The CCMV CP phase diagram established by Bancroft *et al.* has been completed in 2000 by Zlotnick *et al.* [24] who tested assembly at very high protein concentration and observed smaller capsids (pseudo "T=2" particles) of roughly 250 Å.
Lavelle *et al.*, in 2009, [59] also expanded the phase diagram by investigating lower values of pH and higher values of ionic strength. At extreme conditions, they observed novel structures that have not been previously reported, such as disk-like structures and dumbbell-like shells (see Fig 2-14).

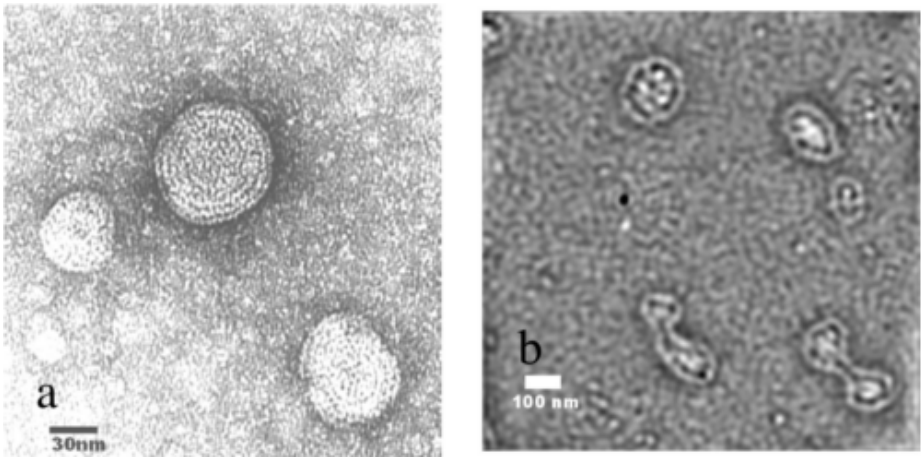


Figure 2-14: TEM images taken from [59]. (a) Disks formed in 0.01 M sodium citrate buffer pH= 4.75, no added salt. (b) Dumbbell structures at pH=7.5, 0.001 M cacodylate buffer, no added salt.

2.3.5 Research highlights on nanoengineering with CCMV

The interest for CCMV has indeed grown because it has shown remarkable properties of assembly. The many prodigious features offered by CCMV have been used at different levels. This section presents shortly the research done on CCMV related to nanoengineering applications.

2.3.5.1 Encapsulation with CCMV coat proteins

The interior cavity of CCMV has been used for the encapsulation of many organic and inorganic compounds including polymers [63], inorganic crystals [64] or nanoemulsions [65]. It should be quoted that there are two approaches to encapsulate materials inside CCMV. The guest molecules can be entrapped during capsid assembly, as it was the case for polymer and nanoemulsion encapsulation, or they can diffuse into the already existing cage architecture through the capsid pores. For the first category, the CCMV coat proteins are in disassembled state and mixed with the nanoparticles

to be encapsulated. The strong electrostatic interactions between the nanoparticles and the CCMV proteins enable the reaction. Douglas and Young [64] reported the encapsulation of polyoxometalate species that belongs to the second category. They controlled the pH-dependent gating of the cages' pores with the pH for that purpose. Empty capsids were first incubated with inorganic tungstate WO_4^{2-} at pH 6.5 and then washed with a solution at pH 5.0 at which the pores closed and the tungstate ions oligomerized form paratungstate ions an crystallize. In a similar way, de la Escosura *et al.* [66] reported the encapsulation of monodisperse prussian blue(PB) nanoparticles inside CCMV capsid. In this case, the crystallization was initiated by UV radiation after diffusion of the PB precursors into the capsid.

2.3.5.2 Functionalization at the interface of CCMV virions

In 2002, the group of Douglas and Young chemically engineered the outer surface of CCMV virions with functional groups to enable site-specific attachment with fluorophores [67]. A range of fluorophore concentrations was used resulting in different degrees of modification (see Fig 2-15). They also have investigated a larger variety

Fluorophore	Mol excess	Dye molecules (av.)
FAM (lysines)	100	180/cage
	300	380/cage
	1000	540/cage
F5M (sulfhydryls)	100	30/cage
	500	60/cage
	1000	100/cage
Cadaverine (carboxylic acids)	100	500/cage
	500	520/cage
	1000	560/cage
SAMSA (bifunctional linker to lysines)	250	45/cage
	500	95/cage
	1000	150/cage

Figure 2-15: Extent of labeling of CCMV using fluorophores that react with different functional groups. A maximum of 540 fluorophores could be attached to the cage. Image taken from [67].

of complex systems to be attached to CCMV exterior surface. Examples include biotin [68], organometallic photosensitizer [69], or even paramagnetic gadolinium atoms Gd^{3+}, used as MRI contrast agents in humans.

The above sections show the versatile nature of CCMV which can combine multiple functionalities at the same time. The possibility to characterize and track those "hybrid" viruses makes CCMV an interesting nanoplatform for biotechnological applications.

Chapter 3

Small-Angle X-ray Scattering (SAXS)

Small-angle scattering (SAS) of X-rays and neutrons has become a powerful technique over the last decades for the study of biological macromolecules in solution. SAS technique allows one to get structural information of biological molecules with relatively high resolution (typically 1 nm). The major advantage of the method is that the system does not need to be ordered, as it is the case for X-ray cristallography. If the information content is not as high as with the latter technique, it is usually sufficient to get global structural information. The biological particles can therefore be studied in solution, allowing the study in their native conditions. This has lead to a new branch of research known as "Bio-SAXS".

This chapter aims at describing shortly the basic physical principles of small-angle scattering, the Time-Resolved Small-Angle X-ray Scattering (TR-SAXS) technique used during this thesis and some data analysis tools based on data fitting that were used to describe the CCMV assembly/disassembly kinetics.

3.1 Basic principles of Small-Angle Scattering (SAS)

This section covers the basic principles of SAS studies. It does not intend to be exhaustive as it is described in more details in most textbooks about scattering techniques (for instance, see references [70], [71] and [72]).

3.1.1 Scattering of X-rays

SAS studies include small-angle scattering of x-rays (SAXS) and of neutrons (SANS). Although the mathematical formalism of x-ray and neutron scattering can be described in the same manner, the physical principles are fundamentally different. This paragraph focuses only on the SAXS method, as it is the main technique used in the thesis.

X-ray photons with an energy E have a wavelength $\lambda = hc/E$ with h the Planck

constant and c the speed of light in vacuum. For structural studies, x-rays with energy around 10 keV are used so that λ is about 0.09-0.15 nm, which is the typical interatomic distance. At this wavelength, the scattering of x-rays by electrons of the atoms can occur.

When an incident plane-wave monochromatic light illuminates the atoms of a molecule with propagation vector $\mathbf{k_i}$ with $|\mathbf{k_i}|=2\pi/\lambda$, it is scattered and $\mathbf{k_s}$ denotes the propagation vector of the scattered light (cf Fig 3-1). We shall consider only elastic scattering (no energy transfer) so that $|\mathbf{k_i}|=|\mathbf{k_s}|=k_0$. The scattering vector $\mathbf{q}$ is defined as the difference between the propagation vectors of the scattered and incident light

$$\begin{aligned} \mathbf{q} &= \mathbf{k_s} - \mathbf{k_i} \\ q = |\mathbf{q}| &= \frac{4\pi}{\lambda} sin\frac{\theta}{2} \end{aligned} \tag{3.1}$$

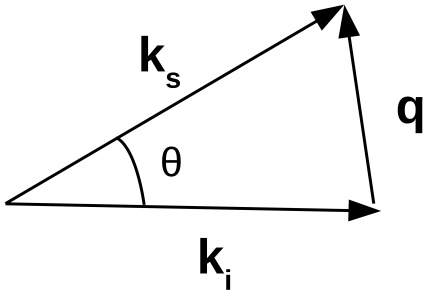


Figure 3-1: Scattering vector $\mathbf{q}$=$\mathbf{k_s}$-$\mathbf{k_i}$.

The scattering process links the real space of coordinates $\mathbf{r}$ and the reciprocal space of coordinates $\mathbf{q}$. In particular, the amplitude of the scattered X-rays by the electrons represents the Fourier transform of the electron density distribution. It is a function of the scattering vectors $\mathbf{q}$ and the scattering length b=$N_e.r_O$, where N_e denotes the number of electrons and r_O=2.82× 10^{-13} cm the Thompson radius.

3.1.2 Scattering by monodisperse particles in solution

In this section, we will consider only monodisperse systems, that is, a solution of N identical particles.

3.1.2.1 Amplitude and intensity of a single particle

The wave is scattered by each atom of the solution which are the ones from the molecules to be studied as well as the ones in the solvent. The scattering experiments therefore implies separate measurements from the solution and the solvent.

It is useful to introduce the scattering length density distribution $\rho(\mathbf{r})$ which is the scattering length distribution divided by the volume V of the solution b($\mathbf{r}$)/V . The difference scattering amplitude from a particle in solution relative to that of the equivalent solvent volume is defined by the Fourier transform of the excess scattering length density $\Delta\rho(\mathbf{r})$=$\rho(\mathbf{r})$-ρ_s, where ρ_s is the scattering length density of the solvent:

$$A_{part}(\mathbf{q}) = FT[\Delta\rho(\mathbf{r})] = \int_{V_{part}} \Delta\rho(\mathbf{r}) e^{i\mathbf{q}.\mathbf{r}} \mathrm{d}\mathbf{r} \tag{3.2}$$

where the integration is made on the volume of the particle V_{part}.

For a given orientation of the particle, the scattered intensity is [73]:

$$I_{part}(\mathbf{q}) = A_{part}(\mathbf{q}) A^*_{part}(\mathbf{q}) = V^2_{part} P_{part}(\mathbf{q}) \tag{3.3}$$

with

$$P_{part}(\mathbf{q}) = \frac{1}{V^2_{part}} \int_{V_{part}} \int_{V_{part}} \Delta\rho(\mathbf{r}) \Delta\rho(\mathbf{r}') e^{i\mathbf{q}.(\mathbf{r}-\mathbf{r}')} \mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \tag{3.4}$$

$P_{part}(\mathbf{q})$ is called the form factor of the particle. It contains the information about how the electrons are distributed within the particle and will be analyzed in more details in following sections.

3.1.2.2 Amplitude and intensity of N identical particles

When the solution is diluted, the particles are widely separated and their positions are not correlated. In that situation, the intensity scattered by the different particles can be added. When N identical particles scatter independently, the total scattered intensity per unit volume V can be expressed as :

$$I(\mathbf{q}) = \frac{N}{V} V^2_{part} P_{part}(\mathbf{q}) \tag{3.5}$$

In a more general case where the particles are correlated in space, the particles are diffusing in time and the intensity is changing for each configuration. The average in time is the measured quantity. For an ergodic system, the intensity is expressed as [73]:

$$I(\mathbf{q}) = \left\langle \frac{A(\mathbf{q}) A^*(\mathbf{q})}{V} \right\rangle_\Omega = \frac{1}{V} \left\langle \left\{ \int_V \Delta\rho(\mathbf{r}) e^{i\mathbf{q}.\mathbf{r}} \mathrm{d}\mathbf{r} \right\} \left\{ \int_V \Delta\rho(\mathbf{r}') e^{i\mathbf{q}.\mathbf{r}'} \mathrm{d}\mathbf{r}' \right\} \right\rangle_\Omega \tag{3.6}$$

where the statistical average $\langle\ \rangle_\Omega$ is taken over the available positions and orientations of the particles. Introducing the center position of a particle $\mathbf{r_i}$ one gets $\mathbf{r} = \mathbf{r_i} + \mathbf{u}$

and

$$I(\mathbf{q}) = \frac{1}{V}\left\langle \left\{ \sum_{i=1}^{N} e^{i\mathbf{q}.\mathbf{r_i}} \int_{Vpart} \Delta\rho(\mathbf{u}) e^{i\mathbf{q}.\mathbf{u}} \mathrm{d}\mathbf{u} \right\} \left\{ \sum_{j=1}^{N} e^{-i\mathbf{q}.\mathbf{r_j}} \int_{Vpart} \Delta\rho(\mathbf{v}) e^{-i\mathbf{q}.\mathbf{v}} \mathrm{d}\mathbf{v} \right\} \right\rangle_{\Omega}$$
$$= \frac{N}{V}\left\langle \left\{ \int_{Vpart} \int_{Vpart} \Delta\rho(\mathbf{u})\Delta\rho(\mathbf{v}) e^{i\mathbf{q}.(\mathbf{u}-\mathbf{v})} \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \right\} \left\{ \frac{1}{N} \sum_{i=1}^{N}\sum_{j=1}^{N} e^{i\mathbf{q}.(\mathbf{r_i}-\mathbf{r_j})} \right\} \right\rangle_{\Omega} \tag{3.7}$$

In the case of an homogeneous particle with excess scattering length density $\overline{\Delta\rho}$, one can rewrite

$$I(\mathbf{q}) = \frac{N}{V}\overline{\Delta\rho}^2 \left\langle \left\{ \int_{Vpart} \int_{Vpart} e^{i\mathbf{q}.(\mathbf{u}-\mathbf{v})} \mathrm{d}\mathbf{u}\mathrm{d}\mathbf{v} \right\} \left\{ \frac{1}{N} \sum_{i=1}^{N}\sum_{j=1}^{N} e^{i\mathbf{q}.(\mathbf{r_i}-\mathbf{r_j})} \right\} \right\rangle_{\Omega} \tag{3.8}$$

When the particles are spherical, the average of the product is in that case equal to the product of the average. Furthermore, in the case of an isotropic system, the intensity only depends on the amplitude of $\mathbf{q}$:

$$I(q) = \frac{N}{V}\overline{\Delta\rho}^2 V_{part}^2 P(q) S(q) \tag{3.9}$$

The first function is the form factor $P(q)$ of the homogeneous single particle arranged such that $P(0) = 1$. The second function $S(q)$ is called the structure factor:

$$S(q) = 1 + \frac{1}{N}\left\langle \sum_{i=1}^{N}\sum_{j\neq i}^{N} e^{i\mathbf{q}.(\mathbf{r_i}-\mathbf{r_j})} \right\rangle_{\Omega} \tag{3.10}$$

The structure factor therefore contains the information on correlation between particles, that is physically the interactions between them. In practice, one can obtain the information on the interactions by first measuring experimentally when the system is diluted. In this case, the expression of the form factor is obtained. The structure factor is then calculated by dividing the measured intensity with the form factor.

In the following, we will consider the ideal case of non-interacting dilute solution of N identical particles, so that the structure factor is neglected and equation 3.5 is applied.

3.1.2.3 Form factor and distance distribution function

Looking back at equation 3.4, averaging over all the orientations of the particles and taking into account that $\langle exp(i\mathbf{q}.\mathbf{r})\rangle_{\Omega} = \frac{sin(qr)}{qr}$, the integration in spherical coordi-

nates leads to :

$$V_{part}^2.P_{part}(q) = 4\pi \int_0^{Dmax} r^2\gamma(r)\frac{sin(qr)}{qr}\mathrm{d}r \tag{3.11}$$

with

$$\gamma(r) = \left\langle \int_{V_{part}} \Delta\rho(\mathbf{r})\Delta\rho(\mathbf{r}+\mathbf{r}')\mathrm{d}\mathbf{r}' \right\rangle_{\Omega} \tag{3.12}$$

being the averaged autocorrelation function of the excess scattering density and D_{max} the maximum particle diameter. In practice, the function $p(r) = r^2\gamma(r)$, called the distance distribution function, is computed with the inverse transformation

$$p(r) = V_{part}^2.\frac{r^2}{2\pi^2}\int_0^{\infty} q^2 P_{part}(q)\frac{sin(qr)}{qr}\mathrm{d}q \tag{3.13}$$

Figure 3-2 displays the form factors (reciprocal space) and the corresponding distance distribution function (real space) of different geometrical bodies with the same D_{max}. Globular particles (red) display a bell-shaped $p(r)$ function with a maximum at about $D_{max}/2$. Elongated particles (green) have a skewed $p(r)$ distribution with a maximum at small distances. Flat particles (yellow) display a broad maximum located at distances smaller than $D_{max}/2$. The $p(r)$ function of hollow particles (blue) has a maximum at large distances. Finally, the $p(r)$ function of particles that consist of multiple subunits that are well-separated display multiple maxima, the first corresponding to the intrasubunit distances and the others to the separation distances between the subunits.

3.1.2.4 Guinier approximation: I_0 and radius of gyration Rg

When $q \to 0$, using a limited development of the eq. 3.11 :

$$\begin{aligned} V_{part}^2.P_{part}(q) &= 4\pi \int_0^{Dmax} \gamma(r)\frac{r}{q}\left[qr - \frac{(qr)^3}{6} + ...\right]\mathrm{d}r \\ &= 4\pi\left[\int_0^{Dmax} r^2\gamma(r)\mathrm{d}r\right]\left[1 - \frac{(qR_g)^2}{3} + ...\right] \end{aligned} \tag{3.14}$$

with

$$R_g^2 = \frac{1}{2}\frac{\int_{V_{part}} r^4\gamma(r)\mathrm{d}r}{\int_{V_{part}} r^2\gamma(r)\mathrm{d}r} \tag{3.15}$$

Using equation (3.14), this yields for the intensity

$$I(q) = I(0)[1 - \frac{1}{3}R_g^2q^2 + O(q^4)] \cong I(0)exp(-\frac{1}{3}R_g^2q^2) \tag{3.16}$$

This is called the Guinier approximation [74] (Guinier & Fournet, 1955) and is

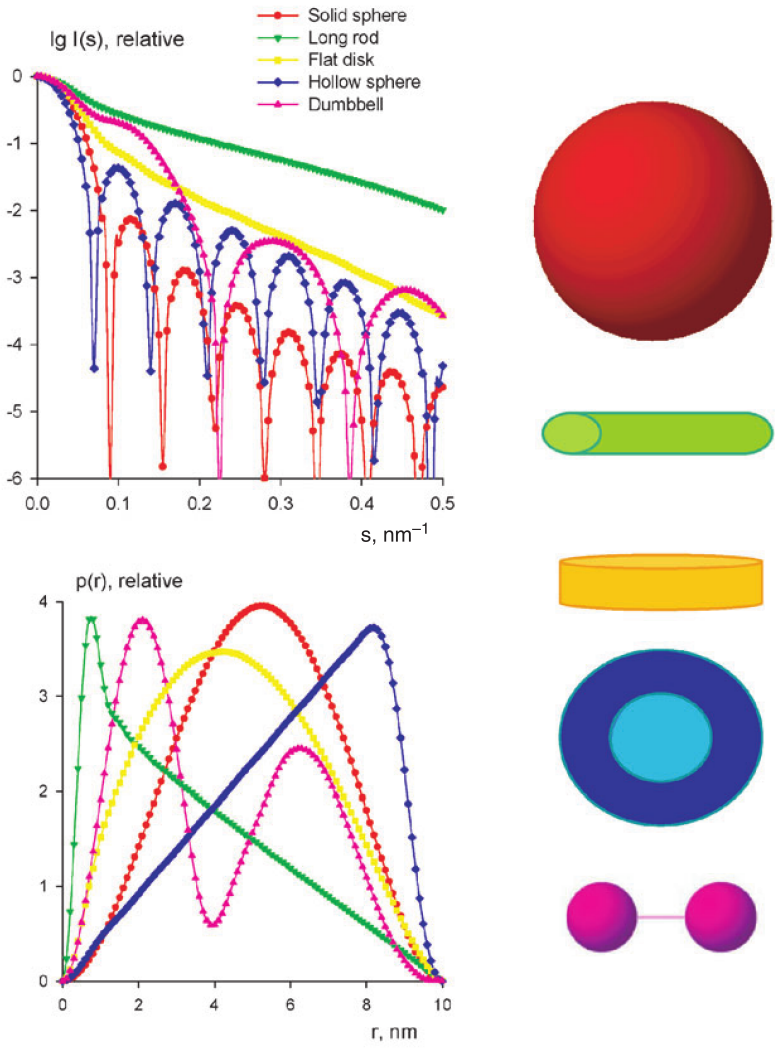


Figure 3-2: Scattering intensities and distance distribution functions of different objects with the same maximum size D_{max}. Adapted from [71].

often the first analysis done with small-angle scattering data. R_g is called the radius of gyration of the particle. It characterizes the size of the particle. I(0) is the intensity at q=0. It is related to the excess scattering length density, the concentration and the volume of the particle. Indeed, in the case of an homogeneous particle with excess scattering length density $\overline{\Delta\rho}$, eq 3.5 gives

$$I(0) = \frac{N}{V}\overline{\Delta\rho}^2 V_{part}^2 \tag{3.17}$$

In practice, I(0) cannot be obtained directly due to technical constraints.The incident X-ray flux is so intense that the detector is masked accordingly. It is rather estimated with the Guinier plot (ln(I(q)) versus q^2). When dealing with absolute intensities (intensity expressed in units of cm^{-1}), one can have access namely to the molar mass and the concentration of the particle. Although it is a powerful technique, one should bear in mind that this approximation is only valid at low angles ($qR_g < 1$).

3.1.2.5 Porod invariant

Using eqs 3.5 and 3.13, one can express

$$\frac{N}{V}\gamma(0) = \frac{1}{2\pi^2}\int_0^\infty q^2 I(q)\mathrm{d}q = \frac{1}{2\pi^2}Q \tag{3.18}$$

In the case of an homogeneous particle, eqs 3.12 and 3.17 lead to

$$\gamma(0) = V_{part}\overline{\Delta\rho}^2 = \frac{I(0)}{\frac{N}{V}V_{part}} \tag{3.19}$$

Combining the two latter equations, one obtains a relationship between Q and the volume of the particle

$$V_{part} = \frac{2\pi^2 I(0)}{Q} \tag{3.20}$$

The integral Q is called the Porod invariant (Porod,1951) since it does not depend on the structure.

3.1.3 Scattering by polydisperse particles in solution

In a dilute system, the average normalized intensity of N identical homogeneous particles in solution expressed in absolute units (cm^{-1}) is (from eq.3.8)

$$I(q) = \frac{N}{V}\overline{\Delta\rho}^2 P(q) = \frac{1}{N_A}\overline{v}^2\overline{\Delta\rho}^2 c^{(m)} M P(q) = K c^{(m)} M P(q) \tag{3.21}$$

where N_A is the Avogadro's number, M the molar mass of the object (in g/mol), $c^{(m)} = NM/VN_A$ its mass concentration (in g/cm^3), $\overline{v}$=VN_A/M its specific volume

(in $cm^3.g^{-1}$), $K = (\overline{v}\Delta\rho)^2/N_A$ (in $mol.g^{-2}.cm^2$) and P(q) the form factor of the homogeneous particle.
When considering particles of different size in solution but with identical excess scattering length density $\overline{\Delta\rho}$ (polymers for instance), the average intensity of the multimeric system in solution (dilute system) at time t can be expressed as

$$I(q,t) = K \sum_i M_i P_i(q) c_i^{(m)}(t) = K \sum_i M_i^2 P_i(q) c_i(t) \tag{3.22}$$

where $P_i, M_i, c_i^{(m)}$ and c_i designate respectively the form factor, the molar mass ($g.mol^{-1}$), the mass concentration ($g.cm^{-3}$) and the molar concentration ($mol.cm^{-3}$) of the oligomer i.

3.2 Time-Resolved Small-Angle X-Ray Scattering (TR-SAXS)

The last decades has seen the rapid improvement of synchrotron techniques developped for real time measurements of rapid kinetics using X-ray and neutron small-angle scattering (SAXS and SANS). These new kind of techniques known as Time-Resolved Small-Angle X-ray/Neutron Scattering (TR-SAXS/TR-SANS), were a major breakthrough for the study of many kinds of kinetics that could occur at the millisecond scale.

3.2.1 Principle

Real time measurements of kinetics often refer to the study of a sample that is subject to a perturbation (dilution, pH jump, temperature change, etc.). The perturbation induces a change in the system which is observed as a function of time. Two main difficulties can be encountered with real time measurements :
- the observation of the very early stages of the kinetics,
- and the time acquisition setup that depends on the acquisition time itself (typically exposure time of the sample to the x-rays/neutrons) and the time between two acquisitions.

The early stages of the kinetics are often difficult to observe. For very short time scales (typically few tens of milliseconds), it is therefore crucial to trigger the acquisition with an automated setup that ensures the difficult coordination between the different devices and also the reproducibility of the experiment. TR-SAXS/SANS experiments thus require high quality apparatus to make possible the observation of rapid kinetics with these constraints. The stopped-flow apparatus emerged as the most frequently used device to study rapid kinetics and will be described in the next

section. The stopped-flow device is coupled with the x-ray/neutron beamline which has been strongly improved over the last decades to enable measurements performance. Studies of millisecond processes now require a high-brilliance X-ray beam combined with a state-of-the-art detector. A second section will describe shortly the synchrotron beamlines where we performed our experiments, i.e on SWING (SOLEIL) and ID02 (ESRF) beamlines.

3.2.2 Stopped-flow device

3.2.2.1 Description

The stopped-flow apparatus allows one to rapidly mix several solutions in a very short time. We used two different stopped-flow devices from the SFM-300 and SFM-400 series (BioLogic, Pont de Claix, France). The SFM-300 (respectively, SFM-400) is composed of 3 (respectively, 4) motorized syringes that are independently controlled and two (respectively, three) mixers as shown in Fig 3-3. Three or four solutions can therefore be mixed at different volume ratios and flow rates. This allows one to better deal with the concentration of the samples as well as its viscosity.

Syringes, delay lines and cuvette

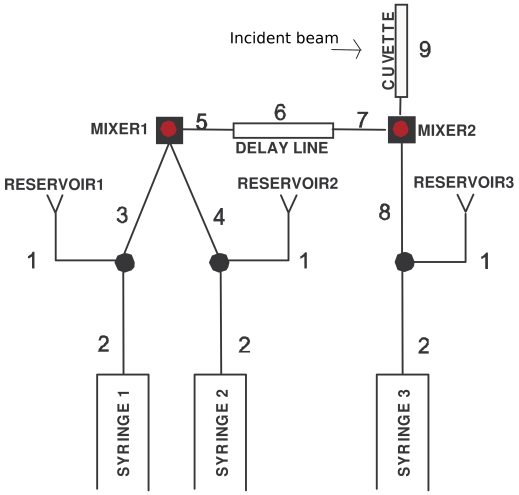
To illustrate an example with the SFM-300 series (Fig 3-3.(a)), the reactants in syringes S1 and S2 are mixed in the first mixer. The mixed solution is again mixed with the solution from S3 in the second mixer and transferred to the cuvette or sample cell (C). The latter is fabricated from a quartz capillary with wall thickness of $\approx$10 μm and diameter $\approx$1.5 mm. The wall thickness is at the same time optimum to withstand the pressure of the flow and to minimize the absorption of x-rays and parasitic scattering effects.

A delay line can be additionally installed by the user. It is an optional tool that aims at varying the ageing time of the solution.

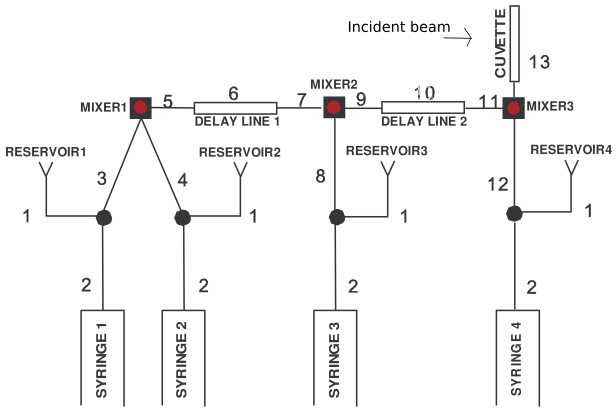
Hard-stop system

During the injection phase, the liquid in the cuvette can reach extremely high velocities and the liquid column must be immobilized in a fraction of milliseconds. With the SFM-300 and SFM-400 series, the flow is immobilized with a high speed electrovalve (hard-stop) which closes the output of the cuvette. Via an adequate synchronization with the motors, the hard-stop system can limit overpressure or underpressure conditions that are potential sources of artifacts.

(a) Picture of a stopped-flow device



(b) SFM-300 schematic



(c) SFM-400 schematic

Figure 3-3: Schematic of stopped-flow devices SFM-300 and SFM-400. The delay lines are optionally added by the experimental user. Adapted from SFM-300/400 Biologic manual.

Mixers and mixing dead time

The mixers can be of two types depending on the SFM series, a standard Berger ball mixer and a high density mixer (model HDS, Biologic) [75]. Both mixers aim at reducing the artefacts from the convection namely due to high differences in solution densities.

The total stopped-flow (mixing) dead time, from which the solution is mixed and is transferred from the mixer to the cuvette, is particularly important to be determined. The dead time is calculated using the known volumes of the cuvette and mixer (dead volume) and the total flow rate.

$$Dead\ Time = \frac{Dead\ volume}{Total\ flow\ rate}$$

Typically, with a SFM-400 with no delay lines, 3 mixers and a total flow rate of 8 mL/s, the stopped-flow dead time is estimated to be 9 ms.

Temperature control

The syringes and the sample cell container are maintained at the desired temperature (between 4 and 70 °C) using an external thermostated circulating bath. The coolant flows through two internal circuits: around the syringes and through the isolation valve block and observation head.

3.2.2.2 Limitations and precautions

This section highlights the limitations and difficulties one can encounter with stopped-flow techniques.

In practice, the reproducibility and the quality of the data essentially depend on the quality of the mixing. The latter is rendered difficult with the difference in densities in the solutions. The flow rates have also to be adjusted accordingly. In particular, cavitation effect can occur at high flow rates (typically >10 mL/s).

Bubble formation also produces artefacts in the data. To limit the presence of bubbles, the solutions have to be degased using vacuum chambers prior to filling the syringes.

Back diffusion of the old sample (sample ejected from syringe) into the fresh sample (sample in the syringe) must be prevented. The use of hard-stop system described earlier considerably limits the risk of such diffusion. However, for more precaution, the syringes are also filled up with the solutions that are chosen accordingly to limit pollution of the syringes.

As a consequence, it is then recommended to test different conditions (by changing the volume ratios and flow rates) for each new system to determine the optimal conditions for the measurement.

3.2.3 X-ray instrumentation for time-resolved experiments

In a time-resolved experiment, the patterns are acquired at closely spaced time intervals and during a time interval that is well-determined. The high brilliance from synchrotron sources now allow to obtain sufficient intensity statistics in an interval of few milliseconds. The synchrotron beamlines are equipped with high sensitivity detectors that are especially required to be fast, i.e having a high count rate (typically detecting $> 10^7$ photons per second)

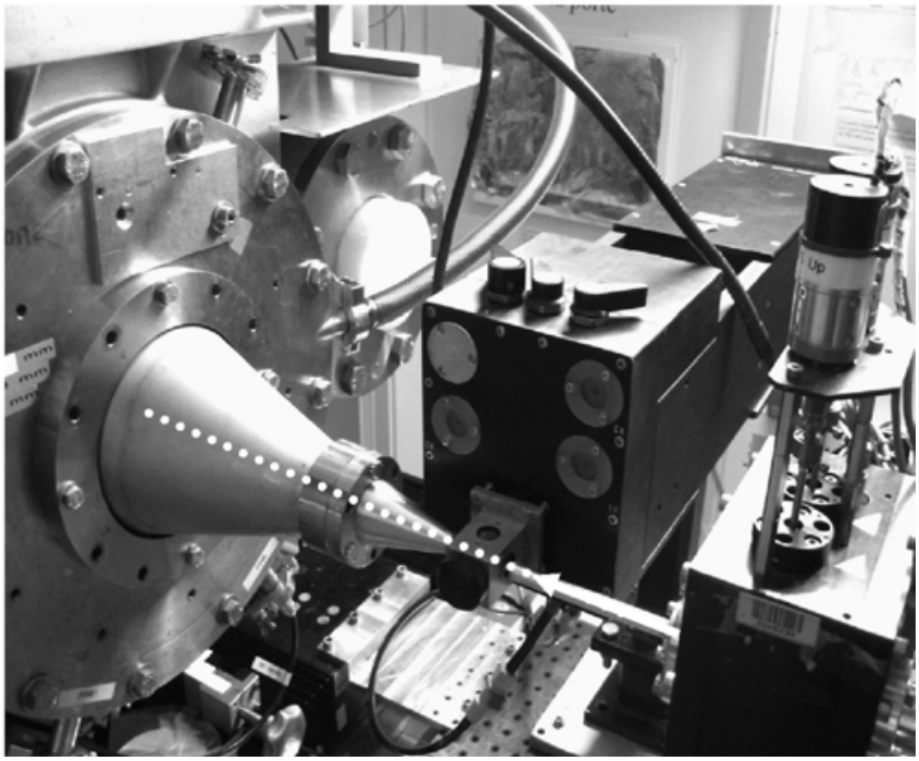


Figure 3-4: Illustration of stopped-flow setup at ID02 beamline (ESRF) with optimized cone for SAXS. White arrow indicates the X-ray beam direction. Image adapted from [75].

ID02 beamline (ESRF)

At the beamline ID02 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France), the incident photon flux on the sample is of about 10^{13} ph/s and the energy is typically 12.5 keV (λ=0.0995 nm). As of September 2014, the main detector used for SAXS studies is the Rayonix MX-170HS and the stopped-flow is the Biologic SFM-400 equipped with four syringes and no delay lines were added.

SWING beamline (SOLEIL)

At the beamline SWING at the SOLEIL synchrotron (Saint-Aubin, France), the energy is typically 16 keV, the detector is AVIEX-PCCD170170 and the equipped stopped-flow is the SFM-300 (three syringes and no delay lines added).

3.2.4 Calibration

3.2.4.1 Absolute units

Absolute calibration of small-angle scattering data refers to calibration of intensities in units of cm^{-1} (absolute units). As seen in section 3.1.3 "Scattering by polydisperse particles in solution" , it is necessary for the determination of molecular weights, volume fraction, the specific volume of the scatterers, etc.

The intensity is usually calibrated through the use of a calibration sample, such as water. The scattering of this sample in absolute intensity units is known beforehand (for water, $\frac{d\Sigma}{d\Omega}$= 0.01633 cm^{-1}). By comparing this known intensity with the intensity as determined on the SAXS instrument, a calibration factor CF can be determined. This is done (with the necessary corrections) through [76, 77]:

$$CF = \left(\frac{d\Sigma}{d\Omega}\right)_{st}(q)\frac{d_{st}T_{st}}{I_{st}(q)} \tag{3.23}$$

where the subscript st denotes the calibration standard, T is the transmission factor, d is the thickness of the sample, I is the measured intensity (after application of the detector and background corrections) and $\left(\frac{d\Sigma}{d\Omega}\right)_{st}(q)$ the known scattering intensity from the calibration sample in cm^{-1}.

The intensity in absolute units of the studied sample is then obtained by multiplying the measured intensity I_{Sample} by CF (following correction by time, sample thickness and transmission).

3.2.4.2 Stopped-flow synchronization

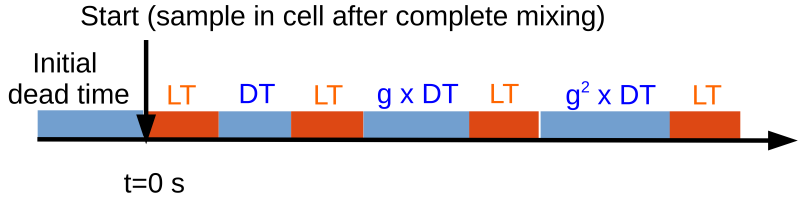


Figure 3-5: Stopped-flow synchronization principle.

The stopped-flow has to be synchronized with the rest of the instrumentation (incident beam, detector,...) in order to avoid to miss the early steps of the kinetics and

to match each of the collected patterns with the corresponding time of the kinetics. The synchronization is done by taking into account the initial dead time (depending on the stopped-flow flow rates and dead volumes and on the detectors), the x-ray exposure time (or life time, LT), the dead time DT between two acquisitions and the geometric factor applied to vary DT.

In the example shown in figure 3-5, the time t is calculated and appended to each spectrum as follows:

$$t(n) = \left(n - \frac{1}{2}\right) \times LT + \frac{1-g^n}{1-g} \times DT, n \geq 1 \tag{3.24}$$

3.3 Mathematical tools for TR-SAXS data analysis

Analyzing the TR-SAXS data often results in fitting the curves with a physical model. This section particularly focuses on some mathematical tools one can use on the TR-SAXS data. The first tool that is introduced is the singular value decomposition (SVD) which is a powerful mathematical analysis that enables to perform data reduction. A second subsection will deal with the global fitting technique and the algorithm developed in this thesis for the purpose of describing the data with a fitting model.

3.3.1 Singular Value Decomposition (SVD)

Practical applications of SVD to data analysis have been found with spectroscopic data for extracting the number of spectral components which are necessary to describe the data and for determining the concentrations of molecular intermediates. Examples of use of SVD analysis concern time-resolved spectroscopic data obtained with absorbance [78, 79, 80] or x-ray scattering [81, 27], the latter technique being considered in the following sections.

3.3.1.1 Mathematical formalism

The measured scattering intensities can be arranged in a matrix $\mathbf{I}$ such that

$$\mathbf{I} = \begin{pmatrix} I(q_1,t_1) & I(q_1,t_2) & \cdots & I(q_1,t_N) \\ I(q_2,t_1) & I(q_2,t_2) & \cdots & I(q_2,t_N) \\ \vdots & \vdots & \ddots & \vdots \\ I(q_M,t_1) & I(q_M,t_2) & \cdots & I(q_M,t_N) \end{pmatrix} \tag{3.25}$$

where q_m and t_n denote the scattering wavenumbers and sampling times respectively.

The matrix of scattering intensities $\mathbf{I}$ of size $M \times N$ $(M > N)$, with M the number of scattering vectors q and N the number of times t, can be expressed as

$\mathbf{I} = \mathbf{USV^T}$ where $\mathbf{U}$ is a $M \times N$ matrix with $\mathbf{U^TU} = \mathbf{I_N}$ ($\mathbf{I_N}$ denotes the $N \times N$ identity matrix), $\mathbf{V}$ is a $N \times N$ matrix with $\mathbf{V^TV} = \mathbf{I_N}$, and $\mathbf{S}$ is a $N \times N$ diagonal matrix of nonnegative elements s_i called singular values of $\mathbf{I}$.

It is important to note that the matrix $\mathbf{V}$ contains the temporal information (in our case, the concentration of the species $c_i(t)$) whereas the columns of $\mathbf{U}$ are eigenvectors that form an orthonormal basis containing the spectral information (in our case, the form factors of the species $P_i(q)$). These eigenvectors have no direct physical meaning and a N-dimensional rotation is required to reconstruct spectra that can be interpretable (see section "SVD rotation procedure").

If the singular values s_i are sorted in decreasing order, the matrices $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{S}$ can be truncated to build a matrix $\tilde{\mathbf{I}} = \mathbf{U_k S_k V_k^T}$ of rank k whose diagonal elements are the k largest singular values of $\mathbf{I}$. The matrix $\tilde{\mathbf{I}}$ provides the best least-squares approximation of rank k of the matrix $\mathbf{I}$. Given the measurement uncertainties, it is possible to describe the scattering intensities with a matrix $\tilde{\mathbf{I}}$ of rank k via the relationship

$$\| \mathbf{I} - \tilde{\mathbf{I}} \|^2 = \sum_{i=k+1}^{N} s_i^2 \leq \mu\nu\sigma^2 \tag{3.26}$$

where σ^2 is a typical variance for the data. The parameters μ and ν can be chosen as $\mu = M - k$ and $\nu = N - k$ [78].

An additional criterion may be required to select components that contain enough information. The relevant information can be quantified by the autocorrelations C of the eigenvectors $\mathbf{U_i}$ and $\mathbf{V_i}$ defined by

$$\begin{aligned} C(\mathbf{U_i}) &= \sum_{j=1}^{M-1} U_{j,i} U_{j+1,i} \\ C(\mathbf{V_i}) &= \sum_{j=1}^{N-1} V_{j,i} V_{j+1,i} \end{aligned} \tag{3.27}$$

Because the eigenvectors are normalized, the autocorrelations close to one indicate a high signal-to-noise ratio and the corresponding eigenvectors can be considered for further processing. Conversely, the autocorrelations that are much less than one, suggest a poor information content and the eigenvectors must be discarded.

Figure 3-6 illustrates an example of SVD analysis performed on TR-SAXS data. These data correspond to the first phase of the disassembly kinetics that will be later investigated in Chapter 4. The graph shows the residual of the truncated matrix of rank k with respect to the initial matrix $\| \mathbf{I} - \tilde{\mathbf{I}} \|^2$ (blue dots) as a function of k. The residual is compared to the experimental noise variance weighted by the number of remaining degrees of freedom $\mu\nu\sigma^2$ (red line). It shows that the matrix of intensities can be truncated to a rank k=2. In other words, this means that the matrix of

intensities can be described by the first two components to within the experimental noise.

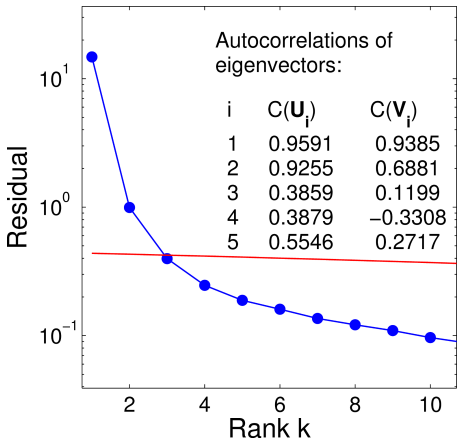


Figure 3-6: Example of application of SVD analysis on TR-SAXS data. The residual $\| \mathbf{I} - \tilde{\mathbf{I}} \|^2$ is plotted as a function of the k largest singular values (blue dots). The residual is compared to the experimental noise variance weighted by the number of remaining degrees of freedom (red line) as explicited by the equation 3.26.

3.3.1.2 SVD rotation procedure

The fact that the matrix of intensities can be described by the first two (or more) components $\mathbf{U_1}$,$\mathbf{U_2}$ (and $\mathbf{S_1}$,$\mathbf{S_2}$,$\mathbf{V_1}$,$\mathbf{V_2}$) does not mean that $\mathbf{U_1}$ and $\mathbf{U_2}$ are the respective species present in the solution. Instead, the signal of these species is somehow "spread" over the first two singular values and vectors.

Here, we limit ourselves to the case of a rank k=2. The signal of the two species $\mathbf{U_{species1}}$,$\mathbf{U_{species2}}$ can be expressed as a linear function of the two components $\mathbf{U_1}$,$\mathbf{U_2}$ that contain all the spectral information from the data. Because $\mathbf{U_1}$,$\mathbf{U_2}$ form an orthonormal basis, it is possible to find θ_1, θ_2 such that

$$\begin{aligned} \mathbf{U_{species1}} &= \mathbf{U_1}.\cos\theta_1 + \mathbf{U_2}.\sin\theta_1 \\ \mathbf{U_{species2}} &= -\mathbf{U_1}.\sin\theta_2 + \mathbf{U_2}.\cos\theta_2 \end{aligned} \tag{3.28}$$

or, in matrix notation

$$\begin{aligned} \begin{pmatrix} \mathbf{U_{species1}} & \mathbf{U_{species2}} \end{pmatrix} &= \begin{pmatrix} \mathbf{U_1} & \mathbf{U_2} \end{pmatrix} \times \begin{pmatrix} \cos\theta_1 & -\sin\theta_2 \\ \sin\theta_1 & \cos\theta_2 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{U_1} & \mathbf{U_2} \end{pmatrix} \times \mathbf{R} \end{aligned} \tag{3.29}$$

$\mathbf{U_{species1}}$ (respectively, $\mathbf{U_{species2}}$) leads to the form factor of species 1 (respectively

2) after normalization. Since $\mathbf{U_k S_k V_k^T}$=$\mathbf{U_k R R^{-1} S_k V_k^T}$ ($\mathbf{U_k}, \mathbf{V_k}, \mathbf{S_k}$ being the truncated matrices at rank k=2), the evolution of the concentration (in arbitrary units) of the species are given by $\mathbf{V_{species1}}$ and $\mathbf{V_{species2}}$

$$\begin{pmatrix} \mathbf{V_{species1}} & \mathbf{V_{species2}} \end{pmatrix} = \mathbf{R^{-1} S_k V_k^T} \tag{3.30}$$

When the evolution of the concentration of species is known, one can have access to the form factors via the SVD analysis. Inversely, the knowledge of the form factors of the species leads to an estimate of the concentration. In practice, of course, nor the concentration neither the form factors are known. Nevertheless, the SVD rotation procedure makes possible to have a good approximation on the calculation of the form factors. Indeed, the form factors are estimated by switching the angles θ_1, θ_2 in such a way that the concentration have a physical meaning (positive values of concentration and evolution in the expected direction). In practice, the range of θ_1, θ_2 that give such physical meaning on the concentration is narrow enough to get a good approximation on the form factors.

Because it is still an approximation that depends on a non-physical parameter (angle θ), the SVD rotation procedure is not used to extract the form factors of the species in the solution, except for the form factor of the capsid of the virus. The latter can indeed be cross-checked with a physical model, which is the form factor of a vesicle model, to ensure they fit at small angles q. The form factor of the capsid that is extracted from the SVD rotation procedure is then preferred to the vesicle model because of the information it provides at higher angles q which is directly obtained from the experimental data.

3.3.2 Global fitting

The term "global fitting" generally refers to simultaneous curve fitting operations performed on multiple datasets. In some cases, the parameters are shared so that a single parameter value applies to all the curves. In some other cases, the parameters are not shared and apply to each curve individually. For each shared parameter, the global fitting procedure yields one (global) best-fit value that applies to all the data sets. For each non-shared parameter, a separate (local) best-fit value for each data set is calculated.

When realizing that the TR-SAXS data are a series of spectra taken at different times in a kinetics, it is clear that in our case, the parameters are shared. These parameters correspond in fact to the rate constants of the kinetics, the total concentration and the stoichiometric coefficients.

In this section, an overview of the global fitting procedure used for the TR-SAXS data analysis will be presented first. Two kind of global fitting algorithms were used in this thesis. The main lines and ideas of these algorithms are exposed in a second

subsection.

3.3.2.1 Overview

In a solution where coexist L different oligomers of the same protein, with no long-range interactions between them, the total scattering intensity is expressed according to equation 3.22, in matrix notation

$$\begin{aligned}\mathbf{I} &= K(\overline{\nu}, \Delta\rho) \begin{pmatrix} M_1^2 P_1(q_1) & \cdots & M_L^2 P_L(q_1) \\ \vdots & \vdots & \vdots \\ M_1^2 P_1(q_M) & \cdots & M_L^2 P_L(q_M) \end{pmatrix} \begin{pmatrix} c_1(t_1) & \cdots & c_1(t_N) \\ \vdots & \vdots & \vdots \\ c_L(t_1) & \cdots & c_L(t_N) \end{pmatrix} \\ &= K(\overline{\nu}, \Delta\rho)\mathbf{P'c}\end{aligned} \tag{3.31}$$

with $K(\overline{\nu}, \Delta\rho)$ a known constant depending upon the specific volume $\overline{\nu}$ and the scattering length density contrast $\Delta\rho$. M_l, P_l and c_l are the molecular weight, the form factor and the molar concentration of species l respectively. The form factor $P_l(q)$ of species l contains the information related to its molecular structure, while $c_l(t)$ specifies the evolution of its concentration over time.

The global fitting procedure consists in finding the form factors P_l and the model parameters $\{\xi_i\}$ embedded into $\mathbf{c}(t, \{\xi_i\})$ such that the relation $\mathbf{I} \approx \mathbf{I^{fit}} = K\mathbf{P'c}(\{\xi_i\})$ is optimal. It can be interpreted using a straightforward least-square approximation which requires that the residual

$$\chi^2 = \frac{1}{MN-1} \sum_{m,n=1}^{M,N} \left(\frac{I_{mn} - I_{mn}^{fit}}{\sigma_{mn}} \right)^2 \tag{3.32}$$

be a minimum with respect to $\mathbf{P'}$ and $\{\xi_i\}$, σ_{mn} being the experimental uncertainties. $\mathbf{P'}$ and $\{\xi_i\}$ are the parameters to be calculated and the global fitting thus consists of a non-linear optimization problem.

The model parameters $\{\xi_i\}$, along with the initial concentrations, govern the evolution of species concentrations through a set of ordinary differential equations:

$$\begin{cases} \dfrac{dc_1}{dt} &= f_1\left(c_1, \cdots, c_L; \{\xi_i\}\right) \\ \vdots & \vdots \qquad\qquad \vdots \\ \dfrac{dc_L}{dt} &= f_L\left(c_1, \cdots, c_L; \{\xi_i\}\right) \end{cases} \tag{3.33}$$

The functions f_i depend on the selected kinetic model. The parameters ξ_i include the stoichiometric coefficients, the reaction rates and the partial reaction orders that depend on the choice of the functions f_i.

3.3.2.2 Global fitting algorithms

Two kinds of global fitting algorithms were developed under Matlab. The first example is an algorithm that determines the whole set of kinetic parameters $\{\xi_i\}$ embedded into $\mathbf{c}(\{\xi_i\})$ as well as the form factors weighted by the molar masses $\mathbf{P'}(q)$.

In the second example, analytical objects are used to model the form factor of the intermediates. These form factors presently depend on few parameters that are related to the geometry of the analytical object in real space. We built the algorithm such that the form factor of the intermediates depend on mainly two parameters which are the number of dimers it is composed of and a typical length that depends on the geometry (for instance $\{\alpha, b\}$). More explanations will be given in Chapter 5 about the model of the intermediate with an analytical object.

Algorithm 1

1. The stoichiometric coefficients to be tested are fixed in the kinetic model.

2. A first loop over the other kinetics parameters $\{\xi_i\}$ (forward and backward reaction rates k_i^+, k_i^- and partial reaction orders μ_i in some cases) is implemented using a constrained nonlinear optimization algorithm. Two such algorithms already exist under Matlab : the Sequential Quadratic Programming (SQP) algorithm [82] and the Interior-point (IP) algorithm [83, 84]. The matrix of concentrations $\mathbf{c}(\{\xi_i\})$ is calculated via the resolution of coupled differential equations as seen in eq 3.33 using the "*ode15s*" implemented Matlab solver. *ode15s* is a solver that namely uses the backward differentiation formulas (BDFs, also known as Gear's method) [85, 86]. It is mainly used to solve stiff differential equations as in our case.

3. Within the first loop, a second optimization algorithm (a simple least-squares algorithm) is run to find $\mathbf{P'}$, such that the relation $\mathbf{I} \approx \mathbf{I^{fit}} = K\mathbf{P'c}(\{\xi_i\})$ is optimal.

4. To find the global minimum, the residual is computed for several initial sets of parameters $\{\xi_i^0\}$ that are randomly chosen (uniform distribution) in their chosen range defined with lower and upper bounds ($\{\xi_i^{min}\}$ and $\{\xi_i^{max}\}$) within the first loop. Typically 30-50 initial points were found to be sufficient to find the global minimum.

   Once the residual χ^2 is found to be a global minimum, its value is stored along with the matrices $\mathbf{c}$ and $\mathbf{P'}$ as well as the set of parameters $\{\xi_i\}$.

5. Steps 2 to 4 are repeated for other sets of stoichiometric coefficients.

Algorithm 2 - use of an analytical object described by the parameters $\{\alpha, b\}$

1. The stoichiometric coefficients to be tested are fixed in the kinetic model. The remaining parameters to be calculated $\{\xi_i\}$ include the kinetics parameters (forward and backward reaction rates k_i^+, k_i^- and partial reaction orders μ_i in some cases) and the geometrical parameters $\{\alpha, b\}$ of the intermediate to be calculated : $\{\xi_i\} = \{k_i^-, k_i^+, \mu_i, \alpha, b\}$.

2. A loop over all the parameters $\{\xi_i\}$ is implemented using a constrained nonlinear optimization algorithm (SQP and IP algorithms). The matrix of concentrations $\mathbf{c}(\{k_i^-, k_i^+, \mu_i\})$ is calculated via the resolution of coupled differential equations as seen in eq 3.33 using the "*ode15s*" implemented Matlab solver. The matrix $\mathbf{P'}$ is computed as a function of $\{\alpha, b\}$.

3. The residual is computed for several initial sets of parameters $\{\xi_i^0\}$ such that $\mathbf{I} \approx \mathbf{I^{fit}} = K\mathbf{P'}(\{\alpha, b\})\mathbf{c}(\{k_i^-, k_i^+, \mu_i\})$ is optimal. The global minimum is searched the same way as described in the algorithm 1. Once the global minimum is found, its value is stored along with the matrices $\mathbf{c}$ and $\mathbf{P'}$ as well as the all set of parameters $\{\xi_i\}$.

4. Steps 2 to 3 are repeated for other sets of stoichiometric coefficients.

Chapter 4

Disassembly kinetics of CCMV empty capsids

If the self-assembly of viruses has been a popular research topic over the last decades, the disassembly kinetics of viruses or virus-like particles has not been as thoroughly studied. One can however wonder if the kinetic pathway of the disassembly process is a mirror image of the self-assembly one. The question about the presence of the same reaction intermediates in these two processes is not at all resolved yet. The way they are interpreted differ from one mechanism to an other. As for the self-assembly, the intermediates often refer to the "nuclei" or clusters of nuclei, which trigger the assembly into complete capsids. As for the disassembly, the intermediates are more seen as some transient states that result from the cracking of the capsid at some specific locations. Yet, the finding of the intermediates in the disassembly process might still reveal some interesting features on the properties of the capsid at a macromolecular level.

We have investigated both assembly and disassembly mechanisms of empty capsids made up of CCMV coat proteins. We decide to begin with this chapter that focuses on the kinetics of the disassembly process since we found the information about the reaction intermediates to be extracted more easily than in the assembly process. The reason for this is that the self-assembly is found to be much faster than the disassembly, at least with the experimental conditions we used. As a result, the lifetime of the intermediates is much shorter, which makes their extraction more difficult.

The CCMV coat protein phase diagram shows stable dimers at pH 7.5 and 0.5 M ionic strength and stable capsids at pH 4.8 and 0.5 M. Therefore, a pH jump from pH 4.8 to pH 7.5 at constant ionic strength (0.5 M) has been chosen to trigger the disassembly.

The first sections deal with the TR-SAXS technique that intends to provide quantitative assessments of the kinetics. The experiments will be presented in sections 4.1,

4.2 and 4.3. A more detailed study involving a model based on fitting will be discussed in section 4.4 and the conclusions will be drawn in section 4.5.

The last section (section 4.6) deals with the study with the light scattering technique. Rather than being quantitative, the study intends to be mainly qualitative. It aims at showing the effect of temperature and concentration on the disassembly kinetics.

4.1 Experiments with TR-SAXS technique

To investigate the disassembly kinetics of CCMV empty capsids, time-resolved small-angle X-ray scattering (TR-SAXS) experiments were carried out with the ID02 beamline at the European Synchrotron Radiation Facility (Grenoble, France) in September 2014. One of the main objectives was indeed to obtain a series of kinetics of disassembly of CCMV empty capsids for quantitative information, which will be presented in section 4.3. The disassembly was chosen to be always triggered with a pH jump from pH 4.8 to pH 7.5 at constant ionic strength (I=0.5 M) but at different temperatures. We also aspired to observe the effect of the pH jump on CCMV virions, which will be presented in section 4.2. The initial concentration of viruses and proteins was not a parameter chosen to be varied since the disassembly process is expected to be a series of first order reactions. The constant rates would not therefore depend on the initial concentration but on the temperature only.

4.1.1 Preparation of samples and buffers

The samples to be studied were :
- a solution of CCMV virions.
- a solution of CCMV coat proteins.

The buffers used in the study of disassembly were :
- the initial buffer (Buffer A): pH 4.8 , 50 mM Sodium Acetate, 0.5 M NaCl, 1 mM EDTA.
- the disassembly buffer (Buffer B): pH 7.5 , 0.55 M Tris-HCl.

An equivalent amount of solution at buffer A and B have to be mixed in order to get a pH of 7.5 and a ionic strength of 0.55 M.

CCMV virions and CCMV coat proteins were purified according to the protocols described in Chapter 2. CCMV virions were dialyzed overnight against buffer A. The concentration was c_{CCMV}=6.5 g/L. The absorbance ratio was measured A_{260}/A_{280} = 1.68. CCMV coat proteins were dialyzed overnight against the same buffer. The concentration was c = 5.8 g/L and the absorbance ratio was verifying $A_{280}/A_{260} \geq$ 1.45.

4.1.2 Experimental setup

4.1.2.1 X-ray and stopped-flow

SAXS patterns were recorded using a high speed and ultra-low noise area CCD detector (Rayonix). The sample-to-detector distance was set to 1.0 m which provided scattering wavenumbers q ($q = 4\pi/\lambda\ sin(\theta/2)$ where θ is the scattering angle and $\lambda =$ 0.996 Å the X-ray wavelength) ranging from 4.2×10^{-3} to 0.39 Å . The experimental setup included a stopped-flow apparatus (Biologic SFM-4000) used to mix the CCMV capsid solutions in acidic buffer with the disassembly buffer B (0.55 M Tris-HCl pH 7.5). The mixed solution was then injected through a quartz capillary of diameter 1.34 mm. The flow was stabilized with a hard stop valve.

The temperature was controlled via an external cryostat (cryo-bath). The cooling was done by circulating a liquid in the stopped-flow device. A temperature probe is in contact with the cuvette to give the precise temperature of the reaction.

4.1.2.2 Data acquisition

The mixing time was 4 ms and the time for the solution to reach the capillary is estimated to be 0.5 ms (initial dead time=cuvette dead volume/total flow rate). Beam exposure time was set to 10 ms. The dead time between two consecutive expositions was set to 33 ms. A geometric factor was applied to the latter to increase the dead time between two exposures as a function of time. The time t appended to each spectrum was calculated as follows:

$$t(1) = \frac{1}{2} \times LT \tag{4.1}$$

$$t(n) = \left(n - \frac{1}{2}\right) \times LT + \frac{1 - g^{n-1}}{1 - g} \times DT, n > 2 \tag{4.2}$$

LT and DT respectively denote the beam exposure time (lifetime) and the time between two expositions (deadtime). g=1.06 is the geometric factor set for the kinetics. The total number of collected frames was chosen to be N=150. With such a setup, t=5 ms (respectively, t=3,244 s) corresponds to the first (respectively, last) acquisition.
The scattering intensities were converted into absolute units after substracting the contribution of the buffer solution. This was simply done by dividing by the diameter of the capillary D=0.134 cm. Experimental uncertainties were calculated after circular averaging using the SAXSutilities software package.

4.2 Presentation of the data - CCMV virions (swelling)

We first intended to probe the effect of a pH jump from pH 4.8 to pH 7.5 at 0.5 M NaCl on CCMV virions. The kinetics was triggered with a pH jump at identical ionic strength and at 10 °C. This section aims at presenting the data collected on CCMV virions only.

4.2.1 Data

Figure 4-1 shows the superimposed spectra of the CCMV virions at pH 4.8 (statics) together with the first and last spectrum of the kinetics (at t=5 ms and t=3244 s). The final concentration was set to be the same between the statics and kinetics experiments, i.e 3.27 g/L.

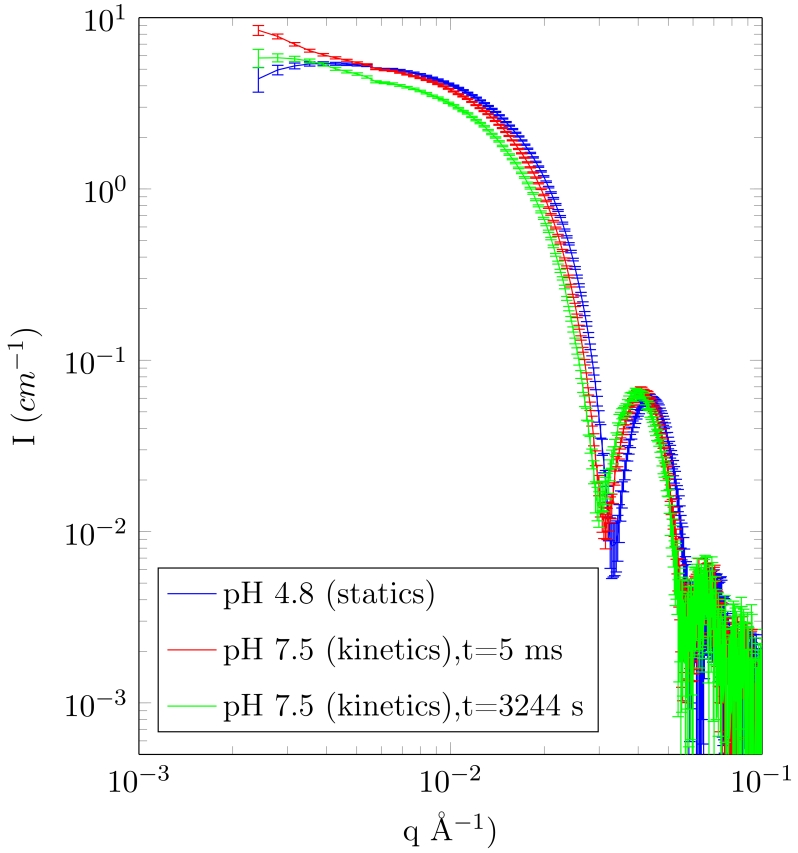


Figure 4-1: Spectra of CCMV virions at different pH at 0.5 M NaCl. The spectrum of CCMV at pH 4.8 (*blue*, static experiment) is compared with the first (*red*) and last (*green*) spectrum of the kinetics triggered by a pH jump from pH 4.8 to pH 7.5. Temperature was 10 °C. Concentration was 3.27 g/L in both experiments.

4.2.2 Data analysis

A small difference in intensity was observed between t=5 ms and t=3244 s. The reason of such a decrease in intensity is not clear but this can be due to the fact that there was probably a small population of empty capsids in the initial solution. This small population of empty capsids may have disassembled giving rise to smaller oligomers, which may explain the decrease in intensity.

The radius of gyration was calculated for each spectrum using the program AUTORG from ATSAS package [87]:
- pH 4.8 (statics): R_g=102.1± 1.5 Å
- pH 7.5 (kinetics) at t=5 ms : R_g=117.6± 2.5 Å
- pH 7.5 (kinetics) at t=3244 s : R_g=116.5± 2.3 Å

The increase in the radius of gyration as the pH increases shows that viruses swell. At pH 7.5, despite the decrease in intensity, the radius of gyration is found to be similar at t=5 ms and at t=3244 s, which supports the fact that the swelling is very fast. This phenomenon has already been reported [16, 88] and the structure of this swollen form of CCMV was determined by cryoTEM and image reconstruction [16]. This has been interpreted as the deprotonation of acidic residues that intensifies the electrostatic repulsion between proteins subunits. The presence of RNA, however, prevents the capsids to disassemble into small oligomers.

With the experimental setup used, one cannot study quantitatively the kinetics of swelling of CCMV virions. Indeed, the first spectrum that can be collected is at t=5 ms and the mixing dead time was 4.5 ms. This means that in the conditions we used (T=10 °C, pH jump from pH 4.8 to pH 7.5), one can only say that it takes less than 10 ms after the beginning of mixing for the CCMV virions to swell.

4.3 Presentation of the data - CCMV empty capsids (disassembly)

In this section, we present the data obtained by TR-SAXS on the disassembly of CCMV empty capsids. It should be reminded that the solution of CCMV empty capsids was obtained by first purifying the CCMV coat proteins from the CCMV virions, and then assembled by dialysis at pH 4.8.

The objectives of the study were to extract the shape of the main intermediates that are showing up in the kinetics and to devise a kinetic scheme that fits the data.

4.3.1 Data

In our experiments, it was decided to keep constant the ionic strength at 0.5 M. The kinetics was always triggered by a pH jump which was from pH 4.8 to pH 7.5. The concentration was not a parameter to vary since the disassembly kinetics can be modeled as a succession of first order reactions, and therefore is expected not to be concentration-dependent:

$$\begin{aligned} \mathrm{C} &\xrightarrow{k_1} \beta_1 \mathrm{I}_1 + \gamma_1 \mathrm{D} \\ \mathrm{I}_1 &\xrightarrow{k_2} \beta_2 \mathrm{I}_2 + \gamma_2 \mathrm{D} \\ &\dots \\ \mathrm{I}_i &\xrightarrow{k_{i+1}} \beta_{\mathrm{i+1}} \mathrm{I}_{i+1} + \gamma_{i+1} D \\ &\dots \\ \mathrm{I}_{N-1} &\xrightarrow{k_N} \gamma_{\mathrm{N}} D \end{aligned}$$

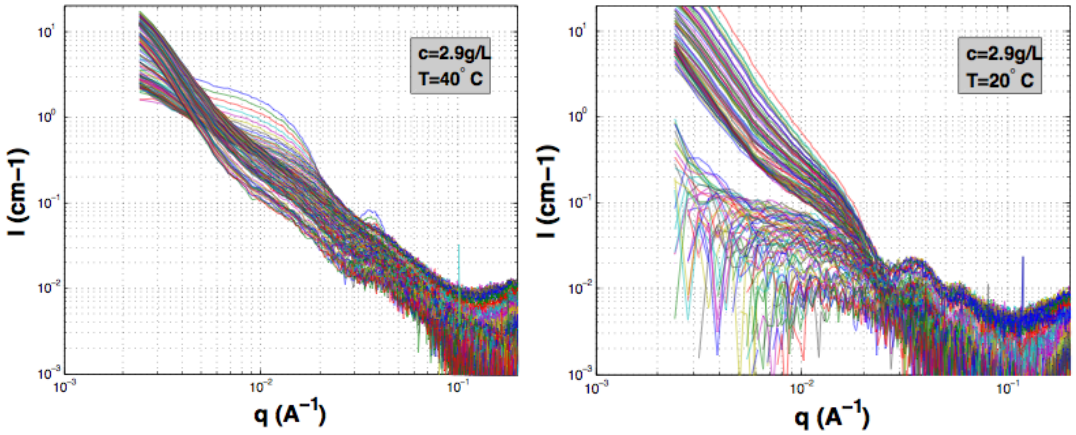
where C symbolizes the capsid, I_i the intermediates and D the dimers. The concentration was however chosen to be relatively high (>2 mg/mL) for a better signal strength and signal-to-noise ratio. This choice was motivated by the fact that the data acquisition setup was set with a short beam exposure time (10 ms). Experiments were done at two temperatures (20 °C and 40 °C). Several attempts have been made to obtain a series of available spectra that could be processed further on. Figure 4-2 shows three different acquisitions at different conditions. Only one out of these three acquisitions can be processed for further quantitative analysis (Fig 4-2.(c)).

4.3.2 First observations

Based on the above data collected at the synchrotron, the following sections focus on the only kinetics available for further processing, which is the one presented in Fig.4-2.(c) where the temperature was T=20°C and the total concentration in proteins c_{prot}=2.9 mg/mL.

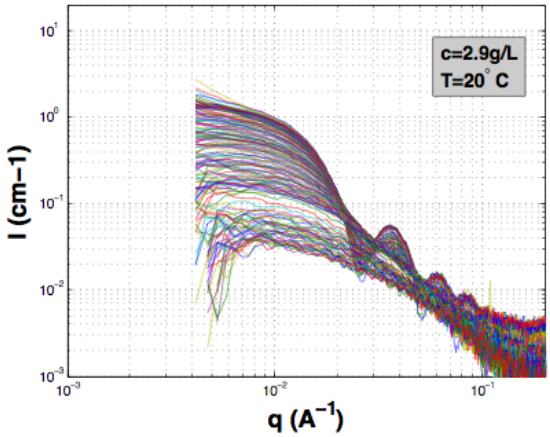
4.3.2.1 Before and after mixing (swelling)

The first spectrum of the disassembly kinetics (at t= 5 ms) is compared with the spectrum at pH 4.8 (statics experiment) that has been collected before the kinetics experiment (Fig.4-3). We observe that just after mixing, the empty capsids start to swell. The observation of swollen empty capsids is new and has not been reported before. Swollen empty capsids are indeed unstable and only swollen virions were previously observed in microscopy (see for example the work of Speir *et al.* [16]). We observed in the former section the swelling of CCMV virions by TR-SAXS. Similarly to the virions, empty capsids swell very quickly, i.e in less than 10 ms, the mixing dead time being 4.5 ms and the time t after mixing being 5 ms.



(a) Strong aggregation of proteins at late times of the kinetics. Data not processed.

(b) Presence of bubbles at early times of the kinetics. Data not processed.



(c) No strong aggregation of proteins nor presence of bubbles at early times of the kinetics. The data can be processed.

Figure 4-2: Examples of collected data on the disassembly kinetics of CCMV empty capsids at different conditions. The main problems faced were the presence of some bubbles at early times of the kinetics just after mixing and the strong aggregation of proteins at 40 °C. Only the kinetics shown in (c) is available for data analysis.

By fitting both patterns with a vesicle model, the outer diameters are found to be 276 Å for the empty capsid at pH 4.8 and 299 Å for the swollen empty capsid.

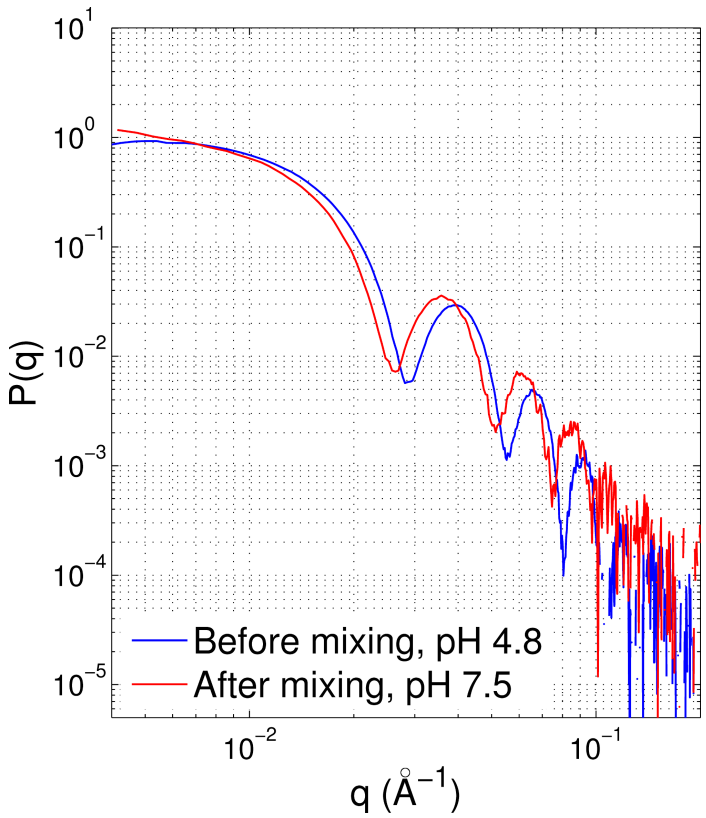


Figure 4-3: Normalized SAXS patterns of CCMV capsid solutions at pH 4.8 (blue curve) (statics) and at pH 7.5 (red curve) $t \sim 5$ ms after mixing (kinetics). The red curve is slightly shifted to the low q values with respect to the blue curve, which indicates that the capsids at pH 7.5 are larger than those at pH 4.8. By fitting both patterns with a vesicle model, the outer diameters are found to be 276 Å(blue curve) and 299 Å(red curve).

4.3.2.2 Evolution of I_0 and isosbestic points

Figure 4-4.(a) shows the time evolution of the forward scattering intensity I_0. It has been calculated by the program AUTORG [87](ATSAS package) available online. The trace displayed a fast step starting from 5 ms and plateauing up to 376 s, followed by a slow step saturating after about 2,000 s (phase III). Within the fast step, we could see a change of the decay rate of $I_0(t)$ at ∼55 s (leftmost arrow on Figure 4-4.(a), which indicated that the fast step could be decomposed into two substeps (phase I and II).

The first and last spectra of the kinetics are also represented in Figs 4-4.(b) and 4-4.(c). Figure 4-4.(b) reveals at least three isosbestic points on superimposed patterns

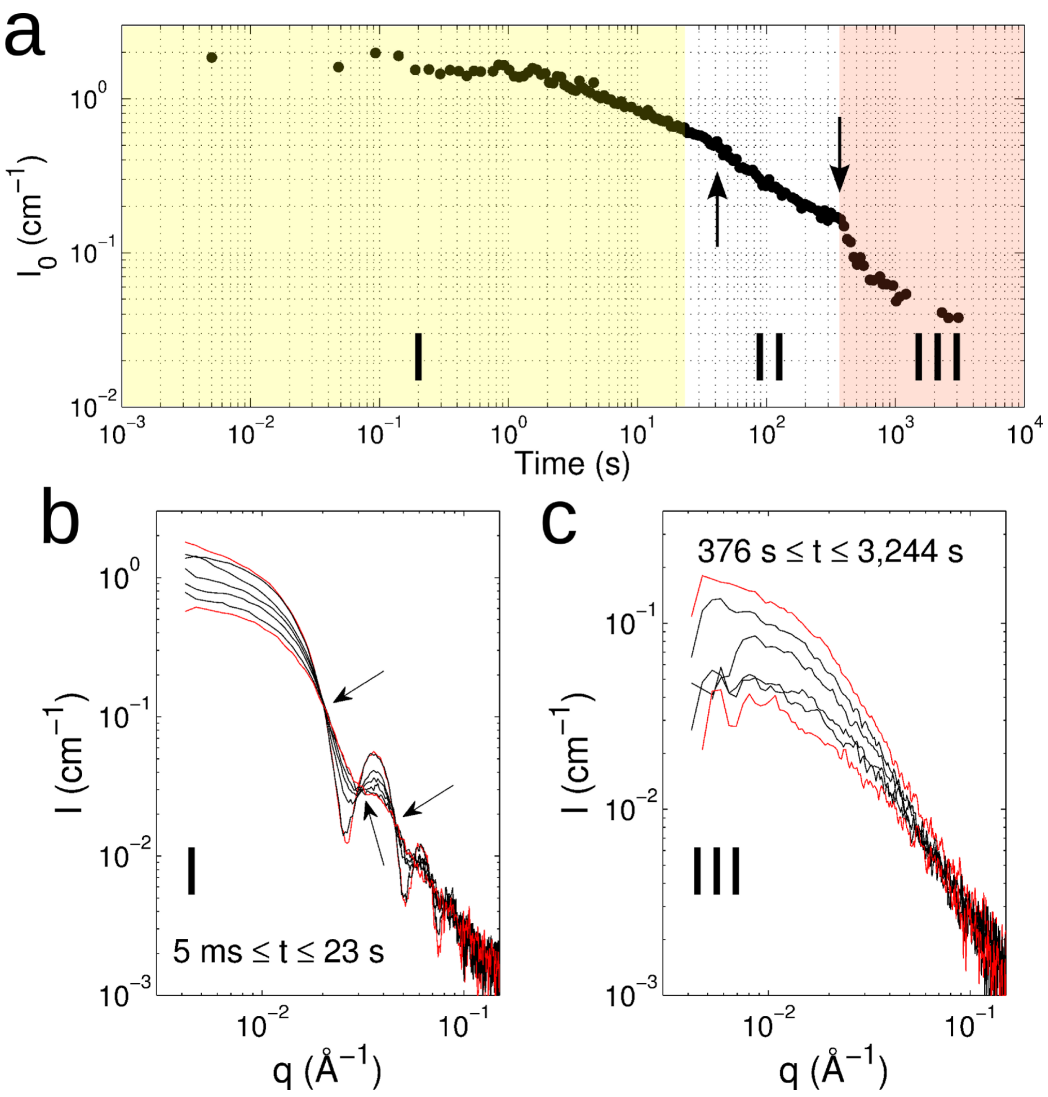


Figure 4-4: TR-SAXS data for the disassembly of CCMV capsids. (a) Time evolution of the forward scattering intensity I_0 from 5 ms to 3,244 s. The phases labelled I, II and III delimit the three phases of the disassembly. The two arrows indicate notable changes in the decay rate of I_0(t). (b,c) Superimposed scattering patterns collected at early (phase I) and late (phase III) times. The patterns in red are the first and the last within each time range. The arrows in (b) point out isosbestic points. The intensities are plotted in absolute units and on a logarithmic scale for better clarity. Only a few patterns are represented also for the sake of clarity. Image adapted from [89].

collected from 5 ms to 23 s and we therefore referred this time range to as phase I. An isosbestic point is a wavenumber at which the scattering intensity does not change during a chemical reaction. Isosbestic points suggested the presence of only two interconverting species that contributed significantly to the intensities [38, 27].

As to the slow step denoted phase III, no isosbestic point could be clearly distinguished (Figure 4-4.(c)). The latest patterns in phase III were identical, and they matched well with the form factors of dimers either measured separately or calculated from the crystal structure (see Figure 4-5).

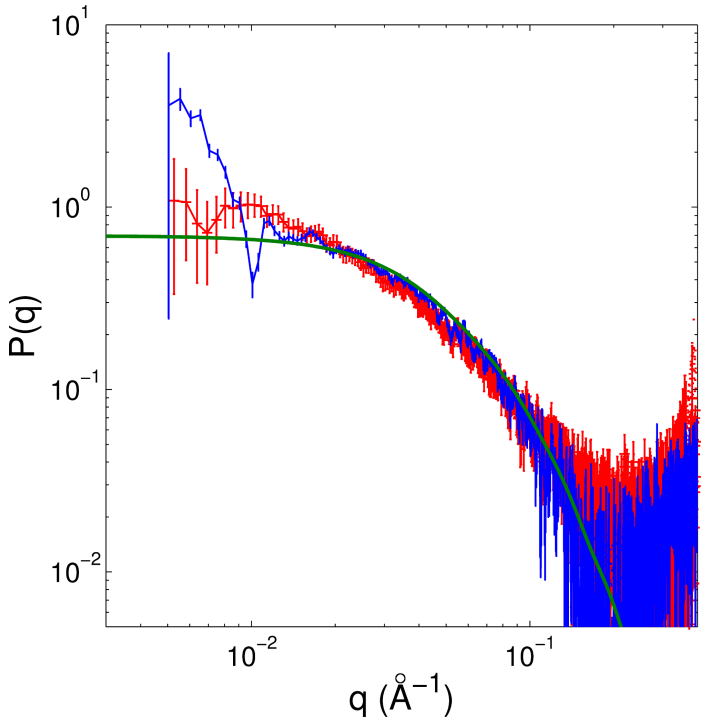


Figure 4-5: Form factors $P(q)$ of CCMV capsid dimers. The red curve is the form factor used in the global fitting. It was obtained by averaging the last three scattering patterns of the data set. The blue curve is the form factor of dimers measured separately in a buffer solution at pH 7.5 and ionic strength 0.5 M, i.e. similar to the ionic conditions of disassembly. The green curve is a form factor calculated by the CRYSOL package [90] from the crystal structure of a CCMV dimer (PDB 1ZA7).

4.3.2.3 SVD analysis

We performed an SVD analysis (see Chapter 3 - Section "Singular Value Decomposition") on the matrix of scattering intensities for early times (phase I), late times (phase III), and the whole set of data.

4.3.2.3.1 SVD analysis - phases I and III - Fig. 4-6.(a) and Fig.4-6.(b)

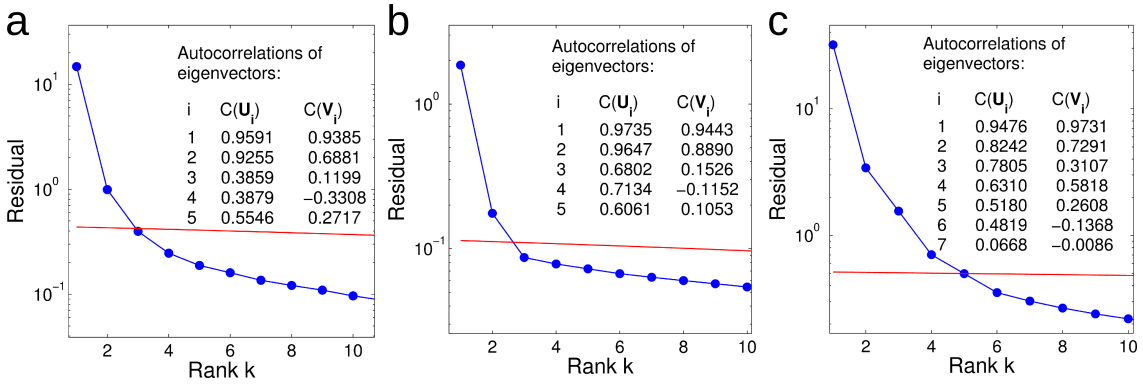


Figure 4-6: SVD analysis of the matrix of scattered intensities for (a) phase I, (b) phase III, and (c) the whole set of data. The graph shows the residual of the truncated matrix of rank k with respect to the initial matrix (blue dots) as explicited in Chapter 3 - Section "Singular Value Decomposition". The residual is compared to the experimental noise variance weighted by the number of remaining degrees of freedom (red line). The values in inset are the autocorrelations of the associated eigenvectors.

For the phase I and phase III, the matrix of intensities $\mathbf{I}$ for 5 ms $\leq t \leq$ 23 s and 376 s $\leq t \leq 3,244$ s could be approximated by a matrix of rank 2 within experimental uncertainties.

For the phase I, the presence of isosbestic points together with the results from the SVD analysis suggest that one can reconstruct the phase I with two major species. Since we had initially capsids noted C, and probably dimers D but with a negligible contribution because of their low molecular weight, capsids were dissociated into a big intermediate I_B during phase I.

For the phase III, no isosbestic point could be clearly distinguished but a SVD analysis also revealed that a matrix of rank 2 was a good approximation of $\mathbf{I}$ in this time range. We could thus deduce that during phase III, that either the big intermediate I_B dissociated into dimers, or it dissociated first into a small intermediate I_S in phase II after dissociated into dimers in phase III.

4.3.2.3.2 SVD analysis - whole set of data - Fig. 4-6.(c)

We then performed a SVD analysis on the whole set of data and we found four components contributing significantly to the intensities, which gave evidence for the presence of four species, that are, C and D, and the two other species I_B and I_S. In conclusion, one can reconstruct the disassembly kinetics with a kinetic model that involves only the form factors of the four species and their concentration as a function of time.

4.4 Modelling the disassembly kinetics of CCMV empty capsids

This section gathers the results from the modelling of the disassembly kinetics of CCMV empty capsids based on a global fitting (see Chapter 3, section "Global fitting" and annex for more details about the model and the algorithm developed for this study). The early and late times (phase I and III) of the kinetics are investigated separately before the study on the whole set of data.

4.4.1 Kinetics at early times (phase I) and late times (phase III)

4.4.1.1 Presence of dimers in the initial solution

The form factor of the capsids was obtained by rotation of the two first SVD eigenvectors on the matrix of data during the phase I. The SVD rotation procedure was explained in chapter 3 section "SVD rotation procedure". The idea was to extract the form factor of the capsids first in order to fix it later in the model and later to extract the form factor of the intermediates. The form factor extracted was consistent with the scattering pattern of a shell with a diameter of 31 nm.

Looking at the forward scattering intensity of the first acquisition (t=5 ms), we noticed that the solution cannot contain only capsids since the intensity I_0 differs from $I = K(\overline{\nu}, \Delta b) \times M_{capsid} c^{(m)}_{protein}$,
where $K(\overline{\nu}, \Delta b)$ a known constant depending upon the specific volume $\overline{\nu}$ and the scattering length density contrast of the subunit Δb, M_{capsid} is the molar mass of the capsid in g/mol and $c^{(m)}_{protein}$ the total concentration in proteins in g/L.

We interpreted this difference with the initial presence of dimers in solution, which is supported by the observation of Zlotnick and coworkers [34, 24]. A fit on the first acquisition (t=5 ms) was therefore carried out with a model of the form

$$I(q) = K(\overline{\nu}, \Delta b) c^{(m)}_{protein} \times [M_{capsid}.P_C(q).x + M_{dimer}.P_D(q).(1-x)] \tag{4.3}$$

The fit (see fig 4-7) shows consistency with the extracted form factor of the capsids and the massic fraction obtained is x=0.5.

4.4.1.2 Global fitting on kinetic data at early times (phase I)

By focusing on the early times of the kinetics (phase I), we aimed at extracting the information concerning the big intermediate I_B. Its form factor $P_B(q)$ was determined with a global fitting on data during phase I, with the following kinetic scheme :
$\mathrm{C} \xrightarrow{k} \beta_B \mathrm{I_B} + \gamma_B \mathrm{D}$
with the mass of the intermediate being such as $M_B = \alpha_B M_D$. The coefficients to be

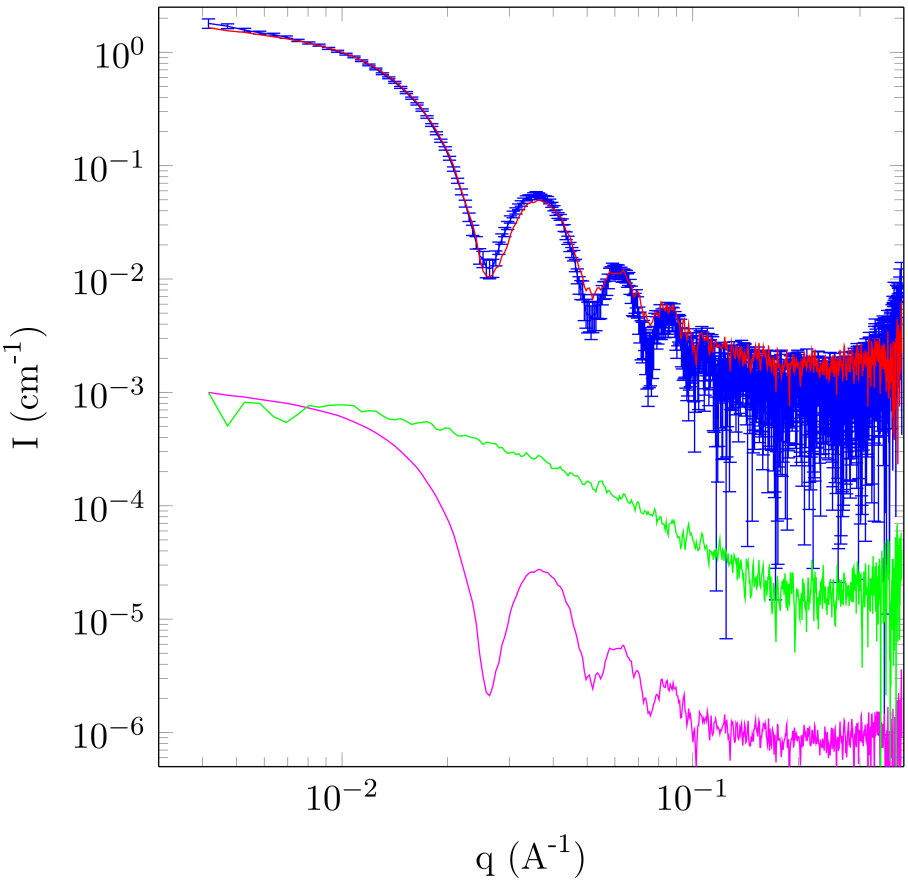
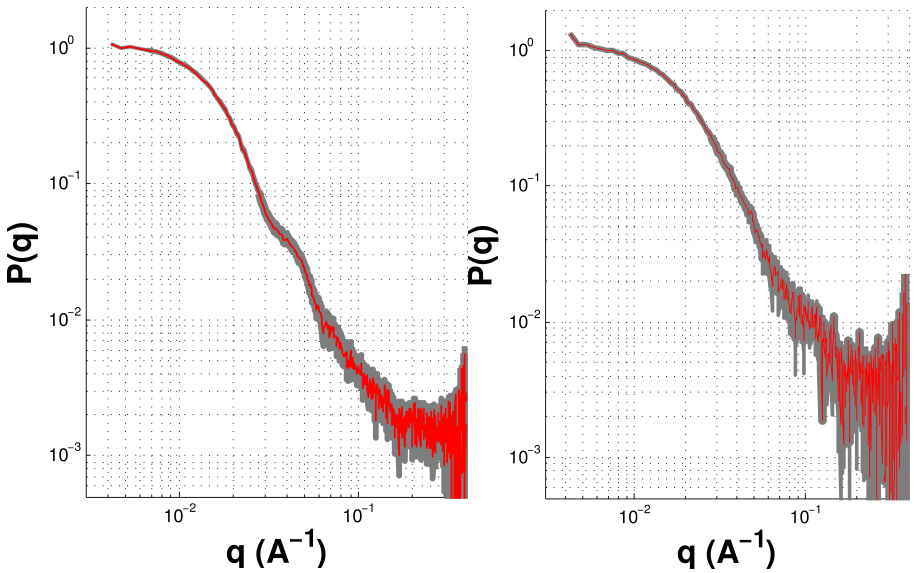


Figure 4-7: Presence of capsids and dimers in the initial solution found at similar mass fraction (1:1). *Blue*: First spectrum of the kinetics (t= 5 ms). *Red*: Sum of the form factors of the capsids (*magenta*, extracted from the SVD analysis at early times of the kinetics) and the dimers (*green*, extracted from the late times of the kinetics) weighted by the concentration (c_{dimers}=$c_{capsids}$=1.45 mg/mL). *Blue* and *red* curves are in units of cm^{-1}, *magenta* and *green* curves are in arbitrary units.

determined were the stoichiometric coefficients β_B and α_B. These were chosen in the range $20 \leq \alpha_B \leq 40$ with $1 \leq \beta_B \leq 3$. In each configuration, the form factor $P_B(q)$ is computed. Figure 4-8.(a) demonstrates that the form factors were robust against the stoichiometric coefficients. This means that the presence of this intermediate is strong during phase I, which is supported by the SVD analysis seen before. The average radius of gyration of I_B was estimated to be 104 Å.



(a) Big intermediate (b) Small intermediate

Figure 4-8: The form factor was extracted by global fitting for a number of meaningful stoichiometric coefficients at (a) early times (phase I) and (b) late times (phase III). The red lines are the averaged form factors and the gray shaded areas represent the range of intensities covered by all the extracted form factors. The radii of gyration were 104 Å and 82 Å respectively.

4.4.1.3 Global fitting on kinetic data at late times (phase III)

The same study was carried out on the kinetics at late times (phase III) with the kinetics scheme : $\mathrm{I_S} \xrightarrow{k} \alpha_S \mathrm{D}$
Even though α_S was varied between 2 and 30, Figure 4-8.(b) shows that the returned form factor of I_S was roughly the same in all cases, leading to the same conclusion about the strong presence of the small intermediate during phase III. The radius of gyration of I_S was found to be 82 Å on average.

4.4.2 Global fitting of the full kinetic data

Based on the above results we carried out a global fitting on the whole set of data with the following scheme

$$\begin{aligned} \mathrm{C} &\xrightarrow{k_1} \beta_\mathrm{B}\mathrm{I_B} + \gamma_\mathrm{B}\mathrm{D} \\ \mathrm{I_B} &\xrightarrow{k_2} \beta_\mathrm{S}\mathrm{I_S} + \gamma_\mathrm{S}\mathrm{D} \\ \mathrm{I_S} &\xrightarrow{k_3} \alpha_\mathrm{S}\mathrm{D} \end{aligned} \tag{4.4}$$

C and D are the capsids and dimers respectively while $\mathrm{I_B}$ and $\mathrm{I_S}$ are the big and small intermediates. β_B, γ_B, β_S, γ_S and α_S are stoichiometric coefficients. Notice that α_S is also the number of dimers making up $\mathrm{I_S}$. k_1, k_2 and k_3 are the forward reaction rates.

C, $\mathrm{I_B}$ and $\mathrm{I_S}$ contained α_C, α_B and α_S dimers D respectively. The choice on the stoichiometric coefficients was restricted with the following conditions

$$\alpha_C = \alpha_B\beta_B + \gamma_B = 90 \tag{4.5}$$

$$\alpha_B = \alpha_S\beta_S + \gamma_S \tag{4.6}$$

The form factors of the four species were fixed in the models and corresponded to the one calculated above since it was shown they were unchanged by the stoichiometric coefficients.

We looked for the set of parameters $\{\alpha_B, \alpha_S, \beta_B, \beta_S, k_1, k_2, k_3\}$ that provided the best approximation of the experimental data in the sense of χ^2 defined by

$$\chi^2 = \frac{1}{MN-1}\sum_{m,n=1}^{M,N}\left(\frac{I_{mn} - I_{mn}^{\mathrm{fit}}}{\sigma_{mn}}\right)^2 \tag{4.7}$$

where $I_{mn} \equiv I(q_m, t_n)$, $I_{mn}^{\mathrm{fit}} \equiv I_{mn}^{\mathrm{fit}}(q_m, t_n, \{\alpha_B, \alpha_S, \beta_B, \beta_S, k_1, k_2, k_3\})$ and σ_{mn} denote the experimental intensities, the fitted intensities and the experimental uncertainties respectively while M and N are the number of scattering vectors q_m and times t_n.

In practice, the set of stoichiometric coefficients $\{\alpha_B, \alpha_S, \beta_B, \beta_S\}$ was fixed but chosen for different relevant configurations, i.e $20 \leq \alpha_B \leq 40$, $1 \leq \beta_B \leq 3$, $2 \leq \alpha_S \leq 20$ and $1 \leq \beta_S \leq 3$ with some forbidden configurations. More details on the algorithm developed for this study are given in Chapter 3 - section "Global fitting" and in annex.

For each configuration tested, we calculated the residual χ^2 as well as the kinetic rate constants. The results for each configuration are shown in figure 4-9. $\beta_B = 1$ and 3 were unlikely values since χ^2 was always greater than 4.0. The most likely pathway, for which χ^2 was below 2.0, corresponded to $\beta_B = 2$ and $\beta_S = 2$, with a big intermediate $\mathrm{I_B}$ made up of $30 \leq \alpha_B \leq 40$ dimers and a small intermediate $\mathrm{I_S}$ made

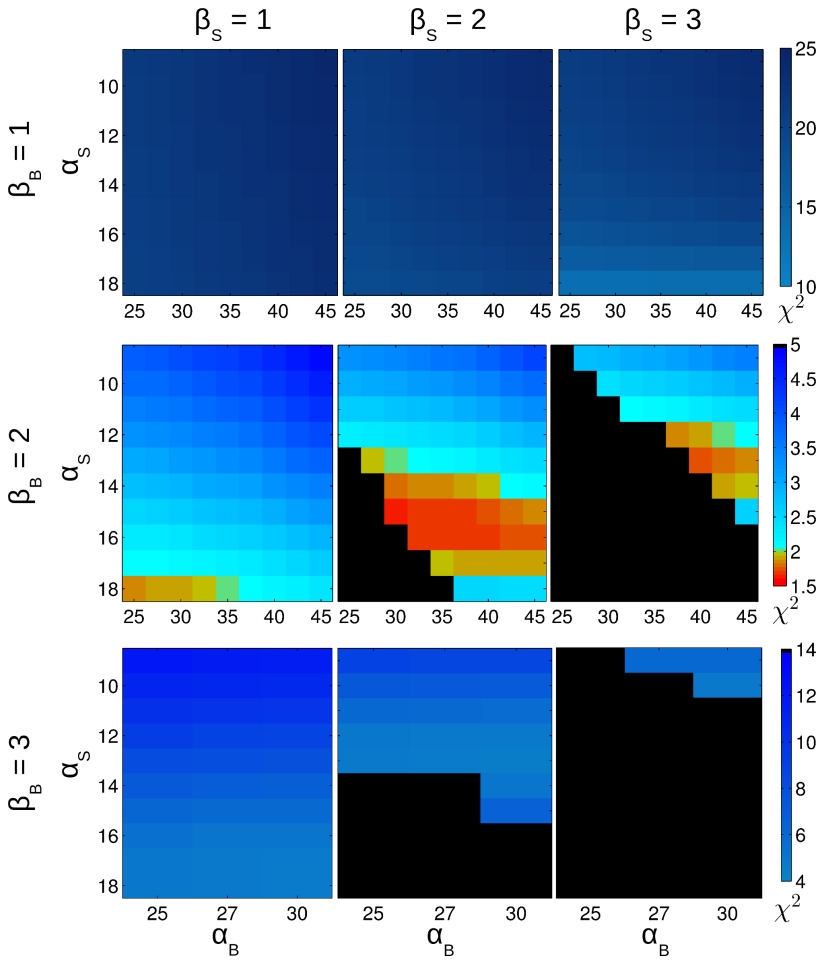


Figure 4-9: Maps of the goodness of fit given by χ^2 for the global fitting of the whole set of data. α_B and α_S are the number of dimers making up I_B and I_S respectively. β_B and β_S are the stoichiometric coefficients entering the kinetic model given by the kinetics scheme 4.4. γ_B and γ_S are computed from Eqs 4.5 and 4.6. Black areas indicate sets of forbidden $\{\alpha_B, \alpha_S\}$ leading to negative values for γ_B and/or γ_S.

up of $14 \leq \alpha_S \leq 17$ dimers.

α_S	α_B	k_1 (s^{-1})	k_2 (s^{-1})	k_3 (s^{-1})	χ^2
14	33	9.28×10^{-2}	8.03×10^{-3}	1.63×10^{-3}	1.83
14	35	9.25×10^{-2}	7.99×10^{-3}	1.63×10^{-3}	1.86
14	37	9.23×10^{-2}	7.96×10^{-3}	1.63×10^{-3}	1.90
15	33	9.20×10^{-2}	9.51×10^{-3}	1.85×10^{-3}	1.67
15	35	9.18×10^{-2}	9.47×10^{-3}	1.85×10^{-3}	1.69
15	37	9.18×10^{-2}	9.42×10^{-3}	1.85×10^{-3}	1.72
16	33	9.18×10^{-2}	1.12×10^{-2}	2.11×10^{-3}	1.68
16	35	9.18×10^{-2}	1.11×10^{-2}	2.12×10^{-3}	1.68
16	37	9.18×10^{-2}	1.11×10^{-2}	2.12×10^{-3}	1.69

Table 4.1: Some reaction rates and χ^2 obtained from global fitting with $\beta_B = 2$ and $\beta_S = 2$.

This finding suggests therefore that capsids first broke into two pieces, which in turn were split into two smaller species before complete dissociation into dimers. We show additional figures to support this result. Table 4.1 gives some reaction rates for stoichiometric coefficients α_B and α_S for which the goodness of fit was acceptable. The calculated reaction rates were similar for each configuration indicating the robustness of the global fitting. Finally, the experimental data are superimposed with the model in Fig.4-10 (top) for β_B=β_S=2, α_B=35 and α_S=16. The corresponding mass fractions of each of the four species are plotted in Fig.4-10 (bottom). Some interesting features of the kinetics can be extrapolated from the traces. After 2,000 s, dimers are the only species present in solution. At 100 s, capsids no longer show up, but the big and small intermediates coexist in equal mass concentrations, with a large population of dimers.

4.4.3 Reconstruction of the intermediates

An *ab initio* shape determination of the two intermediates was carried out with a model of densely packed beads which fitted as closely as possible their extracted form factors (DAMMIN program [91]).

The DAMMIN program consisted of a non-linear optimization solver that used the simulated annealing algorithm. DAMMIN was therefore run twelve times and an average on the twelve structures calculated was done with the program DAMAVER [92] which yielded the most probable model with the lowest NSD (normalized spacial discrepancy) value after refinement. DAMMIN was again run twelve times on the most probable model and the calculated models were averaged. This procedure was done for each of the two extracted form factors.

The resulting low-resolution shapes are shown in Figure 4-11.(b) and 4-11.(c). It can be seen that I_S was elongated (∼220 Å in length) and flat (40-50 Å in thickness),

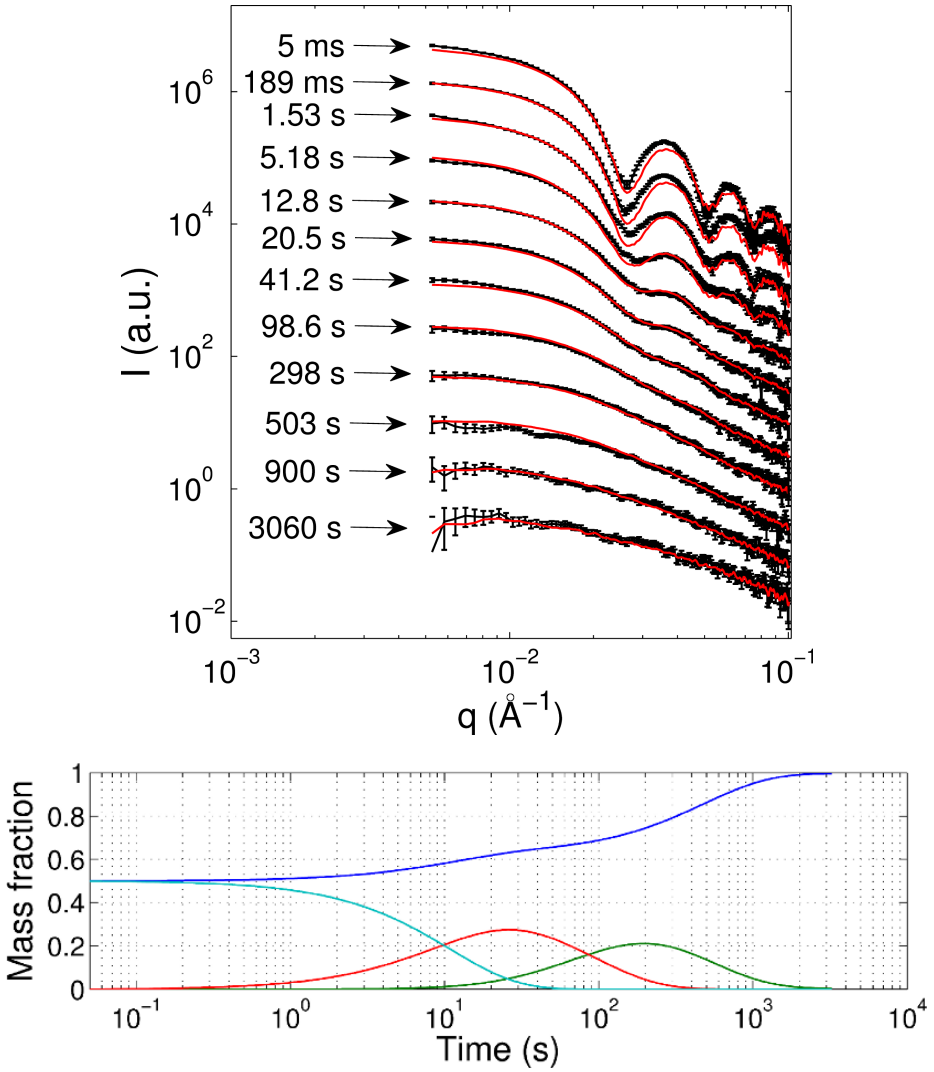


Figure 4-10: *Top*: Comparison between the experimental scattering intensities (black curve) and those reconstructed from global fitting (red curve) at various time points during the disassembly process. The parameters for the kinetic model were set to $\beta_B = \beta_S = 2$, $\alpha_B = 35$, $\alpha_S = 16$, and the resulting goodness of fit was $\chi^2 = 1.68$. The patterns are shifted for the sake of clarity. *Bottom*: The corresponding mass fractions of the four species as a function of time (dimers in blue, small intermediates in green, big intermediates in red and capsids in cyan).

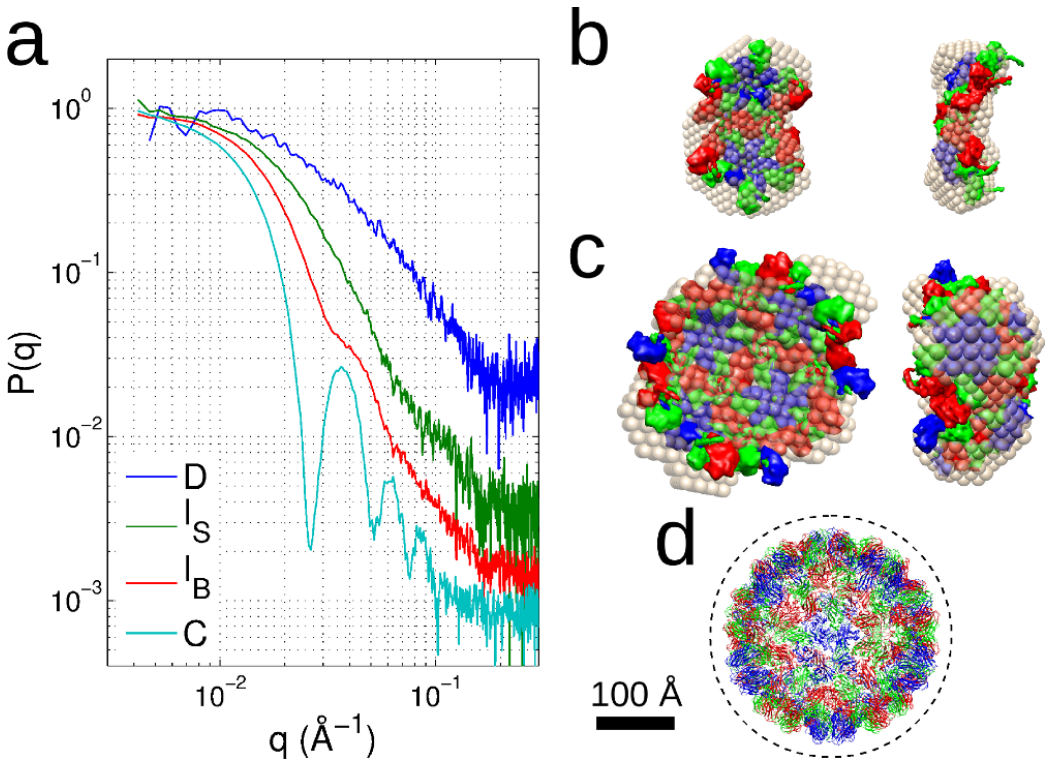


Figure 4-11: Structures of the species involved in the disassembly. (a) Form factors $P(q)$ of the four species in solution: dimers in blue, small intermediates in green, big intermediates in red and capsids in cyan. *Ab initio* shape determinations [91] of the small (b) and big (c) intermediates from their extracted form factors. The superimposed molecular models are from the crystal structure of CCMV capsids (PDB reference 1ZA7). They consist of 15 and 37 dimers for the small (b) and big (c) intermediates respectively. (d) Atomic model of a CCMV capsid in its native (non-swollen) form. The dashed line represents the size of the capsid corresponding to the extracted form factor shown in (a). In all models, proteins 'A', 'B' and 'C' are labelled in blue, green and red respectively.

with a slight curvature. On the other hand, I_B was bulky with approximate dimensions of 330×270 Å and a thickness of 50-60 Å. Next, we extracted from the CCMV crystal structure two putative fragments made of 15 and 37 dimers enclosing two and four fivefold axes respectively. It can be seen in Figures 4-11.(b) and 4-11.(c) that the two molecular models matched quite well the shapes of I_S and I_B respectively. These fragments were good representatives of the actual intermediates with slight variations in the total number of dimers for each structure. One has to keep in mind that these molecular models are directly extracted from the CCMV crystal structure at pH 4.8 (outer diameter of 28 nm). The CCMV empty capsid at pH 7.5 is slightly larger (outer diameter of 31 nm) and the difference is highlighted in figure 4-11.(d).

4.5 Conclusions on the fitting model

Our data were sufficient to detect two major intermediates in the experimental conditions used for that study. The model was also satisfactory when retrieving their average size and shape as well as their typical lifetime. There was not a single set of stoichiometric coefficients but many that give satisfactory results in terms of goodness of fit. A first interpretation would be that this might suggest a multitude of pathways occuring with different probabilities. The form factors of I_B and I_S extracted previously corresponded in fact to averages of structures falling within the above-mentioned range of α_B and α_S. Figure 4-12 illustrates one of the likely pathways: capsids break into two big pieces comprising four or five icosahedral fivefold axes out of the twelves. These big intermediates are in turn dissociated into two smaller intermediates consisting of different sizes, but with an average mass of about 15-16 dimers. The last step is simply the dissociation of these intermediates that do not yield other smaller intermediates with a lifetime long enough to be detected.

Within the experimental conditions, the kinetic model indicated when these intermediates form, coexist and disappear in the kinetics. This may help later to know when to trap them with more sophisticated techniques.

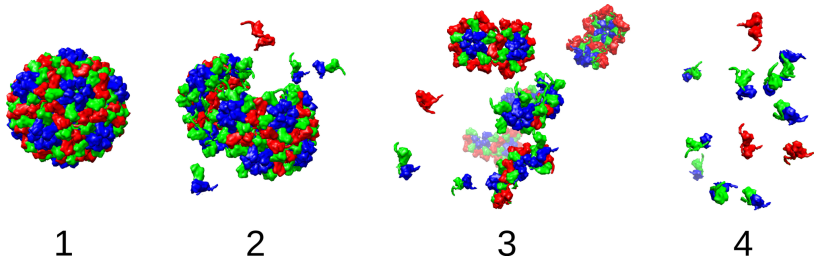


Figure 4-12: A proposed disassembly pathway for CCMV capsid. Proteins 'A', 'B' and 'C' are in blue, green and red respectively. The disassembly steps are labelled from 1 to 4.

4.6 Light scattering experiments

The disassembly kinetics of CCMV empty capsids was also probed by light scattering for complementary observations. With this technique, we intended to visualize the same tendency as with the TR-SAXS results, namely the observation of three different phases. We also carried out more experiments with different conditions of temperature and concentration.

4.6.1 Experimental setup

The solutions were prepared the same way as for the TR-SAXS experiments (purification and preparation of buffers). We used a stopped-flow apparatus (Biologic SFM-3000) supplied with 3 syringes to mix the CCMV capsid solutions with the disassembly buffer B. The mixed solution was injected through a quartz capillary and the flow was stopped by a hard-stop valve. The incident light was produced from a Xe/Hg lamp equipped with an optical filter. The wavelengths tested for the study were $\lambda = 546$ nm and $\lambda = 220$ nm. An external cryo-bath connected to the stopped-flow device was used to control the temperature.

4.6.2 Results and discussion

Figure 4-13 summarizes the results obtained at different temperature and concentration. The reproducibility of the experiment is shown by the superimposition of the two curves at temperature 8 °C and concentration of 1.33 mg/mL. As expected, the disassembly process is not concentration-dependent. As seen before, the disassembly kinetics can be modelled as a succession of first order reactions. On the other hand, the disassembly process is as faster as the temperature is increased.

The kinetics at 20°C displays three regimes such as the kinetics at 20°C observed by TR-SAXS. However, these three phases are shifted. One can estimate roughly three different phases from the slopes (the three rate constants are given in table 4.2). As for a comparison, the first phase approximately occurs from 0 to 5 s (0 to 20 s from TR-SAXS data) and the second phase from 5 s to 100 s (20 to 300 s from TR-SAXS data).

A systematic dramatic increase of the intensity at 1,000 s was observed for each kinetics. It is worth mentioning that the experiments have been done at two different wavelengths ($\lambda = 546$ nm and $\lambda = 220$ nm) which gave the same results. The reason is still unclear. It can be due to technical issues, the hard-stop valve embedded onto the stopped-flow probably failed to stop the flow at this moment. It may also be due to the fact that the proteins aggregate at 1,000 s, which may signify that they are also sensitive to the incident light. This may explain also why the phases are shifted as compared with the TR-SAXS data.

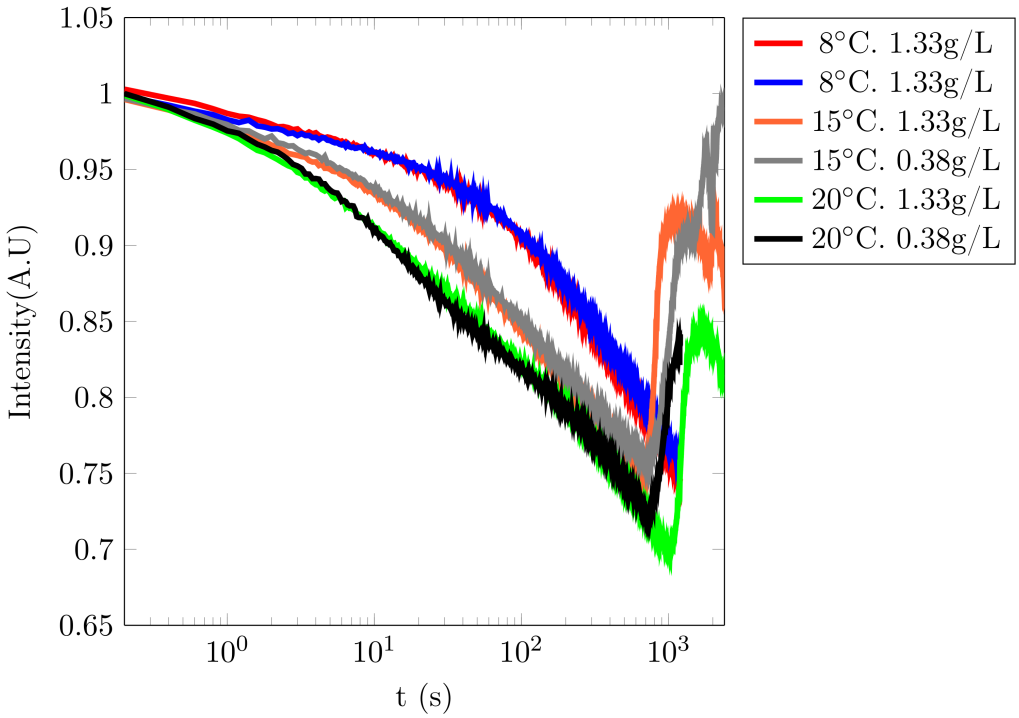


Figure 4-13: Disassembly kinetics of CCMV empty capsids probed by light scattering. The intensities have been normalized to enable comparison with the different kinetics. The wavelength of the incident light is λ = 220 nm and the scattering angle is θ=90°.

Nevertheless, a fit was done on the curves from 0.2 s to 1,000 s with a simplex method with the form:

$$I^{fit}(t) = \sum_{i=1}^{N} c_i e^{-k_i t} + offset \tag{4.8}$$

Two exponential terms (N=2) were not found acceptable for the fit. Three exponential terms (N=3) were thus used to fit all the data.

Table 4.2 gathers the calculated rate constants. By comparing the rate constants with the kinetics at 20 °C calculated from the TR-SAXS data, it is found that the the latter are slightly smaller, meaning that the kinetics observed with light is slightly faster. It is worth noticing that the light scattering experiments have been done twice to check the validity and reproducibility of the data. As for the SAXS experiments, one acquisition was however available. As we switched the temperature at ESRF, it was probably not maintained constant when we carried out our experiments.

Temperature	k_1 (s^{-1})	k_2 (s^{-1})	k_3 (s^{-1})
8 °C	1.6 10^{-1}	1.4 10^{-2}	1.6 10^{-3}
15 °C	1.7 10^{-1}	1.4 10^{-2}	1.0 10^{-3}
20 °C	2.2 10^{-1}	3.9 10^{-2}	1.6 10^{-3}
20 °C (TR-SAXS)	9.2 10^{-2}	1.1 10^{-2}	1.8 10^{-3}

Table 4.2: Measured rate constants for different kinetics at different temperature with light scattering. The table includes the order of magnitude of the rate constants estimated by global fitting with TR-SAXS experiments.

4.7 Conclusions of Chapter 4

In this chapter, we observed the effect of a pH jump from pH 4.8 to pH 7.5 on the CCMV virions as well as on the CCMV empty capsids with the TR-SAXS technique.

In the first case, the effect on the pH made the CCMV virions to swell. This phenomenon has been reported before but the swollen forms were only observed with electron microscopy. With the TR-SAXS technique, it was seen that the swelling process is extremely fast. The transition occurred in less than 10 ms after the beginning of the mixing. The virions were in a swollen form due to the deprotonation of acidic residues which enhanced the electrostatic repulsion between proteins subunits as it was reported in previous investigations [16, 88]. The presence of RNA prevents the virions to break into smaller oligomers due to strong electrostatic interactions between RNA and cationic residues in the interior surface.

This is no longer the case when considering the empty capsids. The latter swell in 5 ms after complete mixing and are not stable. We devised a kinetic model based on global fitting that was robust enough to reconstruct the form factors of two intermediates as well as the kinetic scheme. According to our model, the capsids break into two major pieces that later break into two smaller pieces. The size of these intermediates that we referred to big and small intermediates are made up of about 35 and 16 dimers respectively on average. No more capsids are observed after about 30 s. These intermediates are present from about 1 s to 1,000 s and coexist at equal mass concentration at 100 s. An *ab initio* reconstruction model showed that the extracted form factors were consistent with objects of similar size as calculated in the kinetic model.

The reconstruction of the kinetic pathway was possible with the high spatiotemporal resolution on the ID02 beamline (ESRF). The data contained enough information to extract the intermediates and three phases have been clearly distinguished in the mechanism. These three phases were noted when investigating the disassembly kinetics with light scattering. The latter technique also pointed out the temperature-dependent and the concentration-independent effect on the kinetics.

Chapter 5

Self-assembly kinetics of CCMV empty capsids

The self-assembly of building blocks into larger complex structures is a ubiquitous process in biology. It is especially true when regarding the self-assembly of icosahedral virus capsids, which are designed to encapsulate the genomic material of the virus. In most cases, the self-assembly is both one-component and monodisperse, i.e. the structures formed are made up of multiple copies of the same protein and of a unique and specific size.

While for many cases the processes rely on interactions between the capsid proteins and nucleic acids, capsids have been shown to assemble *in vitro* in the absence of the latter. Therefore, it is primordial to establish a first study on the assembly that consider uniquely the proteins and the interactions between them. The best approach might lie more in the study of the self-assembly kinetics rather than the study at equilibrium. Furthermore, the determination of the nature of the reaction intermediates occurring during the kinetics is particularly challenging. The information one can get from these intermediates are indeed key elements to reconstruct the assembly kinetic pathway.

This chapter is divided as follows :
- The first section of this chapter deals with a literature survey on some works that have already been carried out on the self-assembly of CCMV empty capsids but with different approaches.
- The second and third sections expose the experiments and first analysis we carried out with the TR-SAXS technique. We intended to bring new elements on the CCMV self-assembly by revealing the structures of the intermediates.
- The fourth section is devoted to the modelling of a reaction intermediate with an analytical object used for the reconstruction of the self-assembly kinetics.
- The fifth section presents the results obtained from a global fitting. It shows that the self-assembly kinetics can successfully be reconstructed with a model of an inter-

mediate that has been determined from the data.

5.1 Literature survey on self-assembly of CCMV empty capsids

The nature of the transient intermediates in the process of self-assembly of CCMV empty capsids has been much debated over the last decades. It has been inferred however that the protein subunits should interact and first form a nucleus before completing into a full capsid. This section presents some works that have been carried out on CCMV using different approaches. The studies exposed different arguments to propose a candidate (or several candidates) for the nucleus.

5.1.1 Proposed intermediates from the study of the crystallographic structure

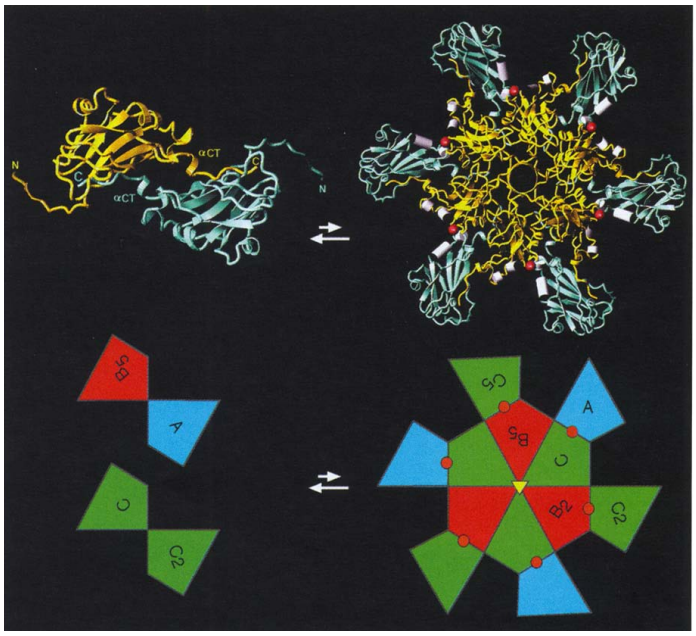


Figure 5-1: Proposed intermediates in CCMV assembly by Speir *et al.* A dimer (left) and a hexamer of dimers (HODs) (right) are displayed both as ribbon diagrams (top) and as polygons (bottom). Color coding of the ribbons is a mix of either B and C, or A and C colors (gold and light blue, respectively). Helices that coordinate calcium (shown as red spheres in the ribbons and polygons) are displayed as white cylinders at subunit interfaces which are also capable of binding RNA. Image taken from [16].

The first proposed intermediates in the assembly of CCMV date back from 1995 when Speir *et al.* resolved the crystallographic structure of the wild-type CCMV

virion [16] and confirmed the T=3 geometry of the CCMV with 12 pentameric and 20 hexameric sites. They proposed that hexamers of dimers (HODs) form the nucleation site for the particle formation (see figure 5-1) given that the dimers (2-mer) and not the monomers (1-mer) are found to coexist in solution. One of their arguments was that the capsid is made up of hexamers of dimers (HODs) and pentamers of dimers (PODs) but HODs have twice the number of interactions per subunit compared with PODs. They completed their argumentation adding that the absence of an ordered amino-terminal structure in the pentameric subunits suggest that they are not formed in solution. It is likely that they only form during the assembly process. Addition of dimers to the HODs would create new interfaces that bind to the RNA and the resulting curvature would drive the formation of the pentamers to form the closed capsid.

5.1.2 Light scattering and pentamers of dimers (PODs)

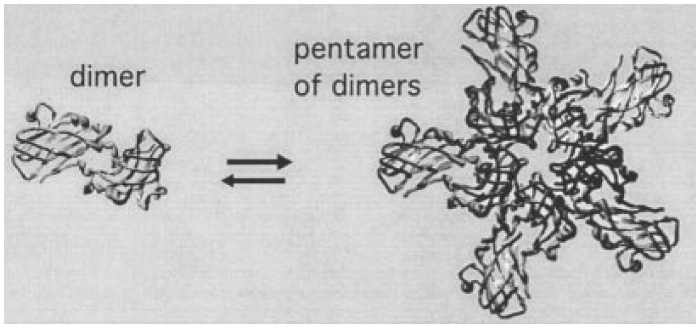


Figure 5-2: Proposed intermediate in CCMV assembly by Zlotnick *et al.* Five dimers (shown left) assemble to form the nucleus, i.e the pentamer of dimers (PODs) (shown right). Image adapted from [24].

In 2000, Zlotnick *et al* [24] studied the kinetics of self-assembly of CCMV empty capsids with the light scattering technique. The kinetics of capsid protein assembly was observed by following changes in intensity of the scattered light. Size-exclusion chromatography (SEC) was also carried out after complete assembly in conjunction with multi-angle light scattering (MALS). Size-exclusion chromatography is a chromatographic method in which molecules in solution are separated by their size and molecular mass. Multi-angle light scattering is a technique for determining the absolute molar mass and the average size of particles in solution. The combined SEC-MALS technique therefore allows the determination of the mass of the reaction products separately.

From the kinetics experiment followed by light scattering and done at different initial concentrations, they estimated the reaction order of the early stages to be five, suggesting that pentamers of dimers (PODs) form at the early stages of the kinetics

(see fig 5-3).
Note: For this study, the pH was 5.25, the ionic strength was 1.0 M and the temperature was 25 °C.

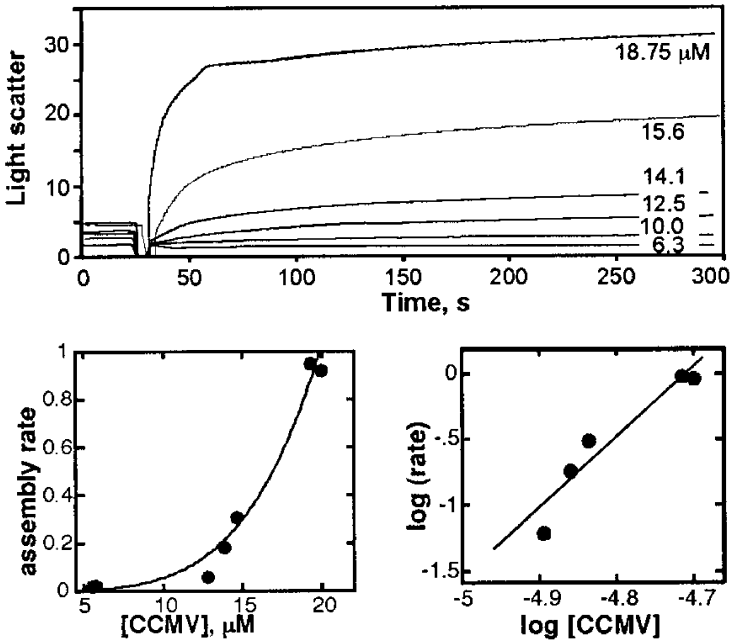


Figure 5-3: CCMV capsid protein assembly observed by light scattering. *Top*: pH 5.25 assembly is concentration dependent. *Bottom left and right*: The order of the fast phase of the pH 5.25 reaction was derived from the concentration dependence of the rate. The least-squares fit of rate versus [subunit] showed a 4.3-power dependence in a linear-linear plot and a 5.5-power law in a log-log plot because the data are weighted differently. Image adapted from [24].

From the combined SEC-MALS technique, the analysis of the chromatograms showed traces of oligomers of 3.5 MDa, 43 kDa and 200 kDa which are consistent with 3.6 MDa capsids, 40 kDa dimers and 200 kDa PODs (see fig.5-4.A). At much higher concentration (fig.5-4.B), the traces of the intermediates were not observed but the chromatograms showed that the capsid peak was shifted to the right. Multi-angle light scattering analysis indicated that the average molecular weight of the peak was about 2.6 MDa. They argued that this was due to a high proportion of 230-Å-diameter particles that form at extreme conditions (high concentrations or pH < 5.0) as seen by electron microscopy. The size and weight were found to be consistent with a 2.4 MDa particle constructed from 60 dimers (a pseudo-T=2 capsid). They later confirmed, the possibility of having such objects in solution with 3D reconstruction from electron microscopy images [17] (see fig.5-5 (Left)).

In parallel, they worked on an assembly model based on the considerations that CCMV empty capsids can form a T=3 capsid and a pseudo T=2 capsid at higher

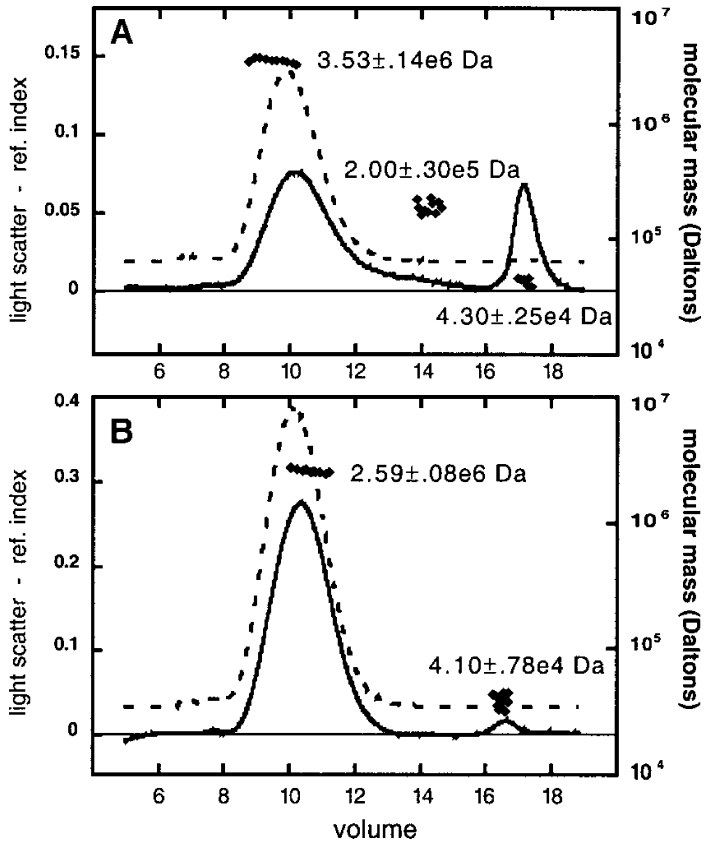


Figure 5-4: Analysis of capsid assembly by light scattering and size-exclusion chromatography at pH 5.25 and I=1.0 M. Elution volume in x-axis is in mL. (A) For 22 μM capsid protein, major assembly products were 3.5 MDa T=3 capsid and 43 kDa dimer. The right edge of the capsid peak was polydisperse and trailed off to a 200-kDa species (i.e., a pentamer of dimers). (B) Assembly of 53 μM capsid protein yielded peaks with estimated molecular weights of 2.6 MDa and 41 kDa. The capsid peak was shifted 0.4 ml compared to a T=3 capsid due to the presence of pseudo T=2 capsids. No intermediate species were observed. Chromatograms show the refractive index (solid line), 90°light scattering (dashed line), and molecular weight estimated from Debye plots (diamonds). Image adapted from [24].

concentration [34] (see fig.5-5 (Right)). They proposed that the formation of 90-dimer (T=3) particles begins with the formation of PODs. The T=3 capsid is then formed by the addition of subsequent dimers and PODs. A shunt was also included for formation of 60-dimer (pseudo T=2) particles when there are high concentrations of PODs.

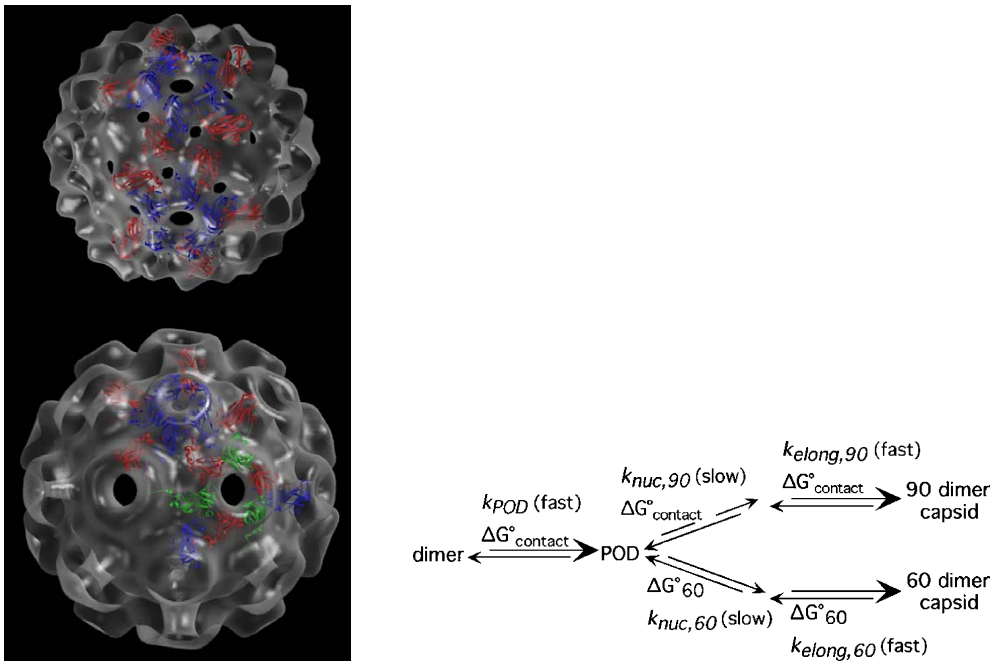


Figure 5-5: *Left*: Cryo-EM reconstruction images from electron microscopy of a pseudo T=2 capsid (top) and a T=3 capsid (bottom). Color coding refers to the usual notations for the capsid (proteins A in blue, B in red and C in green). Image adapted from [17]. *Right*: Assembly model proposed by Zlotnick *et al* [34]. The POD is the assembly nucleus on which dimers and then other PODs will aggregate and form the closed capsid.

5.1.3 Computational analysis - Energetics and structural characteristics of a viral capsid

Theoretical studies have also been carried out to elucidate the assembly pathways of capsids. One of them focuses on the information obtained from the structure and the organization of the subunits in the complete capsid. These information come from the nature and the position of the atoms constituting the capsid calculated from x-ray crystallography and stored in a PDB file. Electrostatic and van der Waals contributions to the protein interface energies can be computed, giving more information on the capsid structure and stability.

Reddy *et al.* [93, 94] addressed the energetic characteristics in the protein capsid and utilized the information at the protein interfaces to compute and compare potential assembly pathways. Briefly, they first identified the building blocks (protein subunits) of the assembly looking at the stronger association energies they computed from the structures. Then, they used a combinatorial approach to determine potential intermediates that may be formed during the capsid formation. A substructure of n subunits is considered a likely intermediate when the association/binding energy of formation of the substructure is favored significantly over any other combinations of n subunit associations. Finally, they looked at the most likely substructures to identify the most likely pathway of capsid formation with the following argument that the stronger the protein-protein interactions, the greater the chance of determining the intermediates of the capsid assembly pathway.

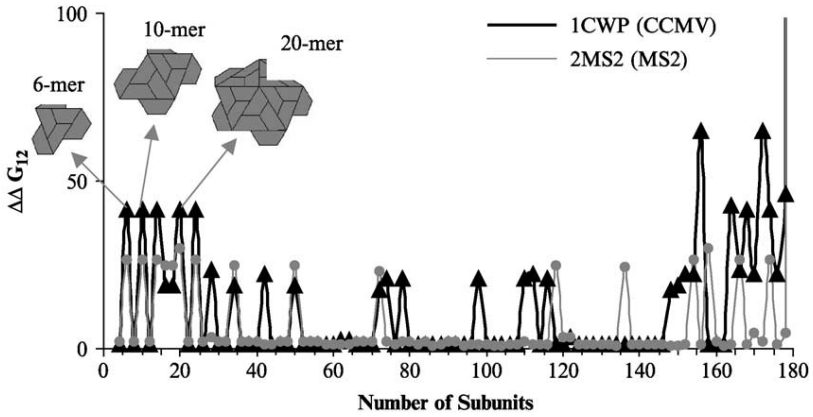


Figure 5-6: Assembly profile of CCMV (PDB ID: 1CWP) (thick black lines) that was calculated by the addition of subunit dimers. The results suggest that the 6-mer is the nucleus on which dimers will add to form other intermediates before completing the capsid. Reddy *et al.* also studied the assembly of other viruses such as the bacteriophage MS2 (PDB ID: 2MS2) (thin gray line). Image adapted from [94].

Figure 5-6 illustrates the assembly profile of CCMV obtained from these energetics calculations. The assembly pathway suggests that the first preferred structure is a trimer of dimers (6-mer), followed by a pentamer of dimers (10-mer) and a decamer of dimers (20-mer) surrounding the fivefold axes of symmetry. One should notice that the preferred decamer in their analysis is different from that reported by Zlotnick and coworkers. They concluded that the nucleating structure is the 6-mer (trimer of dimers) followed by the addition of dimers to arrange a pentameric vertex representing a decamer of dimers (20-mer). Further addition of dimers complete the formation of the capsid.

5.1.4 Conclusions and discussion

The last three subsections have highlighted that the assembly pathway of CCMV is still unsolved and still much debated. Different approaches were used to try to put forward some arguments that may bring elements to understand the self-assembly mechanism.

So far, the assembly model of Reddy and coworkers is a predictive model although they extracted numerous information from the CCMV capsid structure. Furthermore, their calculations were only based on association/binding energies of a structure of n subunits from the preferred structures of the *n-1* subunits pre-calculated. Therefore, their model excluded from the start the possibility of a cooperative assembly between oligomers.

From the experimental results of Zlotnick and coworkers, the detection of pentamers of dimers (PODs) in solution has been put forward. The combined SEC-MALS experiment carried out separately showed that capsids elute at 10 mL, dimers at 17 mL and some oligomers elute at 14 mL (fig 5-4). The SEC technique aims at separating the oligomers as a function of their molecular weight. They were thus able to measure the (average) mass of the oligomers that eluted on the chromatogram, provided their concentration was high enough to have a sufficient signal.

However, the study of assembly with light scattering was found less convincing. In particular, the method yielding the power/rate law from the kinetics at different initial concentrations seemed unappropriated. They indeed based their calculations on the determination of the initial rate which has been calculated from the first points of the kinetics. This was prejudicial because oligomers of higher mass than a POD could readily coexist during the first seconds of the kinetics. As a consequence, their assumption of having a unique nucleus to determine its size was from the start disputable. The calculated power law (five) would rather resulted from the fact that it was an average value.

5.2 TR-SAXS experiments and data acquisition

TR-SAXS experiments were carried out on the self-assembly of CCMV empty capsids at the SOLEIL synchrotron (SWING beamline) in June 2012. The combined stopped-flow and x-ray scattering was a method of choice for this study. On the first hand, the SAXS experiments aimed at collecting spectral information which is not provided by light scattering. On the other hand, the stopped-flow device is designed to study rapid kinetics. This apparatus can trigger a rapid mixing at several volume ratios and flow rates to follow the very early steps of the reaction. The stopped-flow apparatus was essential for these experiments since the self-assembly kinetics of CCMV capsids was chosen to be triggered with a pH jump and different initial protein concentrations were investigated.

5.2.1 Preparation of samples and buffers

Two buffers were used for the experiment :
- the sample buffer (Buffer A): pH 7.5, 50 mM Tris-HCl, 0.5 M NaCl, 1 mM EDTA.
- the assembly buffer (Buffer B): pH 4.8, 50 mM sodium acetate, 0.5 M NaCl, 1 mM EDTA.

CCMV coat proteins were purified according to the protocols described in Chapter 2. A day before the experiments, CCMV coat proteins were dialysed overnight against the buffer A. The absorbance ratio and the concentration of the sample were verified with absorbance measurements.

5.2.2 Experimental setup

5.2.2.1 X-ray and stopped-flow

SAXS patterns were recorded using a CCD detector (AVIEX). The sample-to-detector distance was set to 4.94 m. With this distance, the range of scattering vectors q was between 8.9 10^{-3} Å^{-1} and 0.5 Å^{-1}. The stopped-flow apparatus was the Biologic SFM-300 which is equipped with 3 syringes. The temperature was maintained at 10°C with a thermostated circulating water bath.

5.2.2.2 Data acquisition

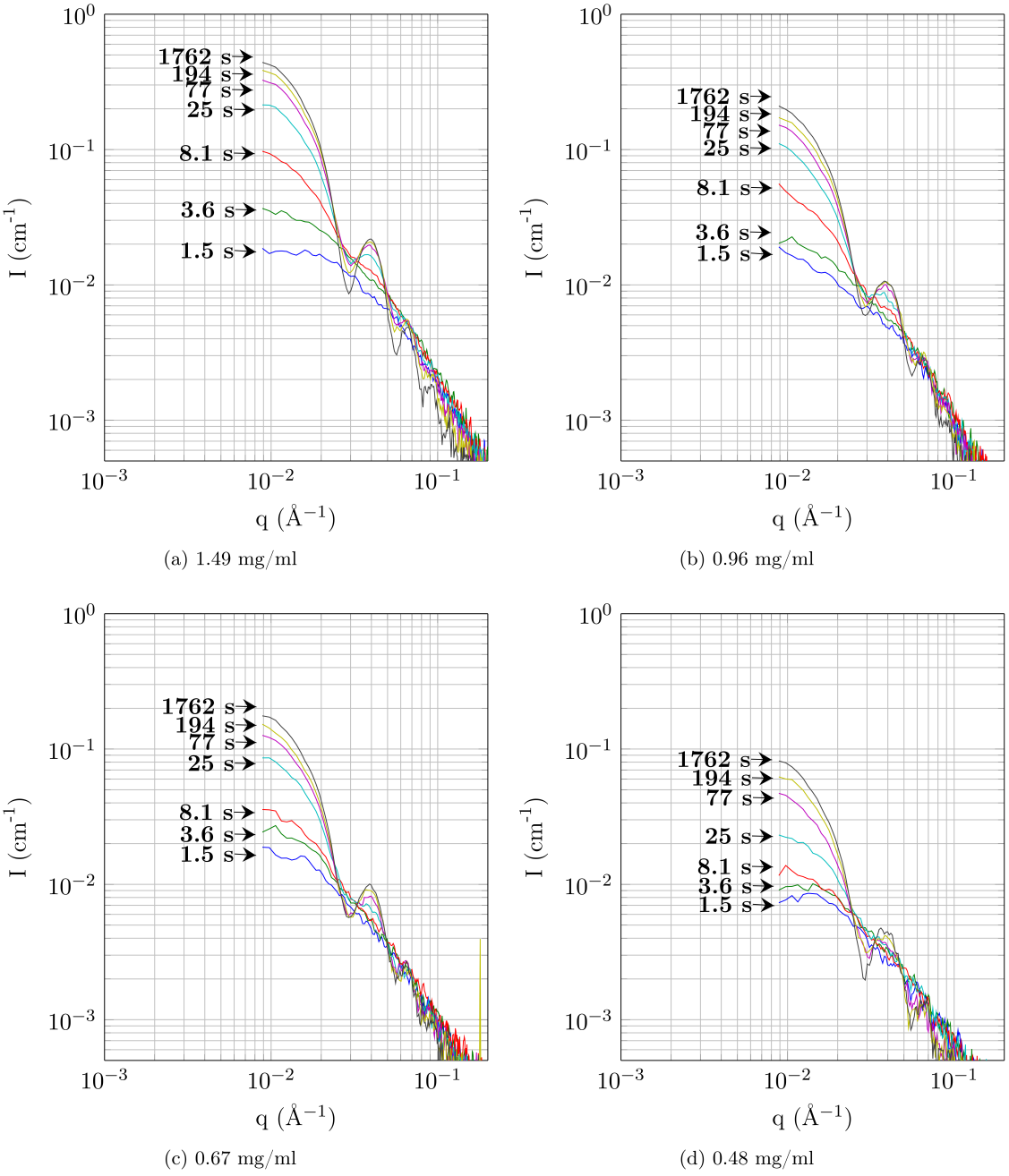
We calculated the time t that we appended to each spectrum as follows:

$$t(n) = \left(n - \frac{1}{2}\right) \times LT + \frac{1-g^n}{1-g} \times DT, n \geq 1 \tag{5.1}$$

LT=1 s and DT=1 s respectively denote the beam exposure time (lifetime) and the time between two expositions (deadtime). g=1.1 is the geometric factor set for the kinetics. The total number of collected frames was N=54. With such a setup, t= 1.5 s (respectively, t=1762 s) corresponds to the first (respectively, last) acquisition. The scattering intensities were converted into absolute units after substracting the contribution of the buffer solution. Experimental uncertainties were calculated after circular averaging using the "Foxtrot" software, which is the software used on the SWING beamline to process the data.

5.2.3 Data and reproducibility

Four series of kinetic data at four different initial protein concentration were collected and processed with the "Foxtrot" software. The four concentrations were 1.49



(a) 1.49 mg/ml

(b) 0.96 mg/ml

(c) 0.67 mg/ml

(d) 0.48 mg/ml

Figure 5-7: Self-assembly kinetics of CCMV empty capsids by TR-SAXS at four different concentrations. The data were collected on the SWING beamline at the SOLEIL synchrotron in June 2012.

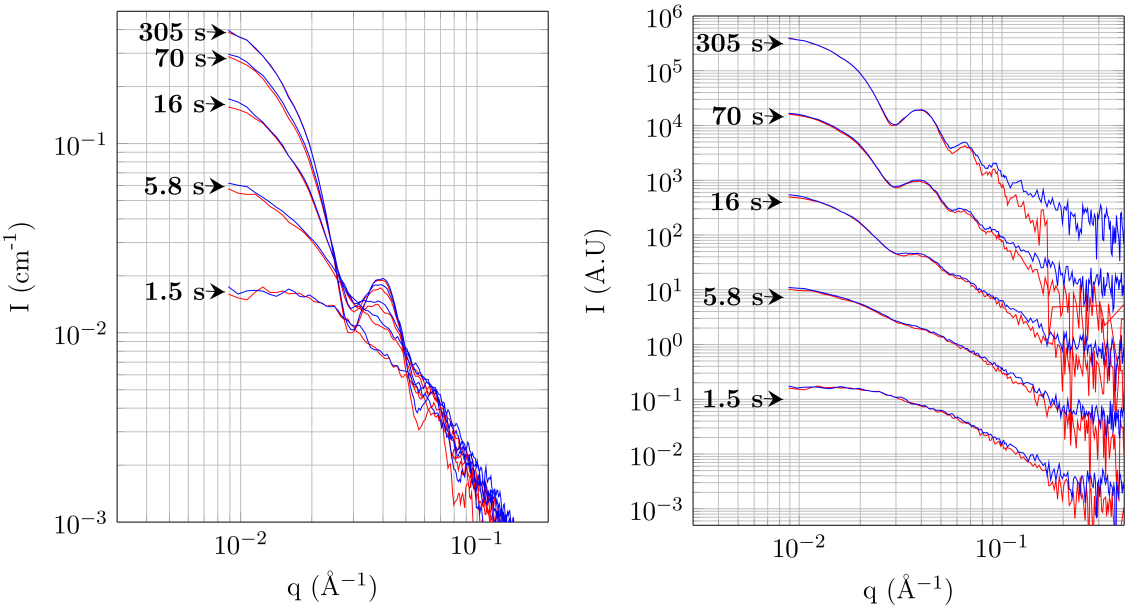


Figure 5-8: Reproducibility of the TR-SAXS experiments. *Left* : The two collected kinetic data series at 1.49 mg/mL superimpose very well at small angles. *Right*: Same spectra but shifted for clarity. The x-axis is extended to 0.4 Å^{-1} to illustrate the difference in intensities at higher angles due to the substraction of the buffer. One of the kinetic acquisition ends at 305 s.

mg/mL, 0.96 mg/mL, 0.67 mg/mL and 0.48 mg/mL. The data are shown on figure 5-7.

A second kinetic data series at 1.49 mg/mL was collected from t=1.5 s to t=305 s. Figure 5-8 illustrates the high reproducibility of the experiment. The patterns superimpose very well at small angles. There is a difference in intensities at higher angles (q>0.1-0.2 Å^{-1}) due to some imperfections resulting from the substraction of the buffer. The reproducibility showed that the mixing with the stopped-flow was found to be adequate. No bubbles were formed after 1.5 s with this mixing at 1.49 mg/mL.

5.3 First analysis on TR-SAXS data

When dealing with such data, the first concern is to know from which material do the capsids start to assemble, i.e, what are the protein subunits at the beginning of the kinetics (monomer, dimer or trimer, etc.). The second information that is needed is the number of species that are showing up and that are predominant in the kinetics. A kinetics model can then be devised according to these information.

5.3.1 Dimers

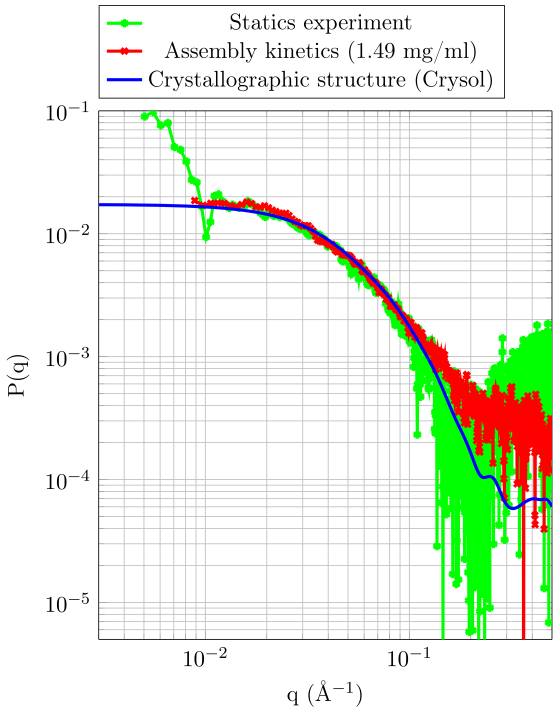


Figure 5-9: Form factors P(q) of CCMV capsid dimers. The red curve is the first spectrum of the assembly kinetics at 1.49 mg/mL. The green curve is the form factor of dimers measured separately in a buffer solution at pH 7.5 and ionic strength 0.5 M. The blue curve is a form factor calculated by the CRYSOL package [90] from the crystal structure of a CCMV dimer (PDB 1ZA7). Errorbars are removed for the sake of clarity.

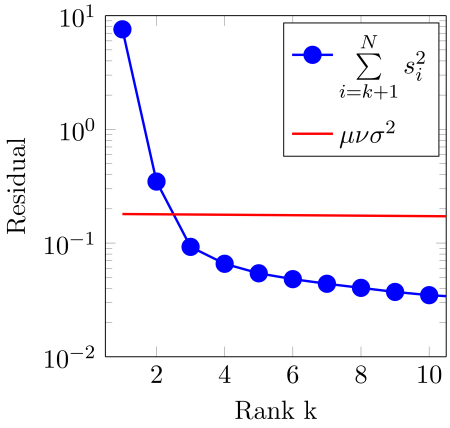
Adolph and Butler [52] , and later Zlotnick *et al.* [24] observed that the protein subunits at pH 7.5 and ionic strength 0.5 M are dimers, respectively by analytical ultracentrifugation and size-exclusion chromatography coupled with multi-angle light scattering.

We checked again this finding by SAXS. Figure 5-9 shows the superimposition of different spectra of dimers. The green curve is obtained by measuring in static condition in a buffer solution at pH 7.5 and ionic strength 0.5 M. The blue curve is the theoretical spectrum of the dimers calculated from the crystallographic structure (PDB 1ZA7) with the CRYSOL package [90]. The red curve corresponds to the first acquisition of the assembly kinetics, after mixing to a final pH 5.25 and final ionic strength I=0.5 M. The first acquisition is at t=1.5 s meaning that the major oligomers are under dimers form.

5.3.2 SVD analysis

A SVD analysis was performed on the whole set of data by concatenating the four matrices of concentrations $I^{(k)}(q,t)$ (with k=1,2,3,4 representing the four different concentrations) into a single matrix $I_{cat}(q,t)$.

$$I_{cat}(q,t) = \left[\ I^{(1)}(q,t)\ \middle|\ I^{(2)}(q,t)\ \middle|\ I^{(3)}(q,t)\ \middle|\ I^{(4)}(q,t)\ \right] \tag{5.2}$$



i	$C(\mathbf{U_i})$	$C(\mathbf{V_i})$
1	0.9121	0.9869
2	0.9751	0.9424
3	0.4529	0.1289
4	0.5189	0.3492
5	0.4765	0.1145
6	0.2055	0.0484

Figure 5-10: SVD analysis on the concatenated matrix. *Left* : Residual of the truncated matrix of rank k with respect to the initial matrix (blue dots). The residual is compared to the experimental noise variance weighted by the number of remaining degrees of freedom (red line). *Right* : Autocorrelations of the associated eigenvectors.

The results are illustrated in figure 5-10. It shows that only two components contribute to the signal. Their signals are strong enough (high signal-to-noise ratio) so that the intensities can be described by these two components. Since dimers are present at the beginning of the kinetics, and capsids at the end, one would say that only dimers and capsids contribute to the signal. However, an increase in mass from the very first spectra of each kinetics is observed, and these spectra do not present oscillations typical from the capsid. This means that there is at least one intermediate that is showing up. The two components from the SVD analysis are therefore rather assimilated as the signals from the capsid and one intermediate since the signal of larger species could predominate the signal from the dimers.

5.4 Modeling the reaction intermediate with an analytical object

The unknown reaction intermediate is modeled as an object that can be expressed analytically. The mass/volume and dimensions are important parameters to be deter-

mined (fit parameters). This section shows how the analytical object is constructed and the derivation of its form factor.

5.4.1 Dimensions of the empty capsids

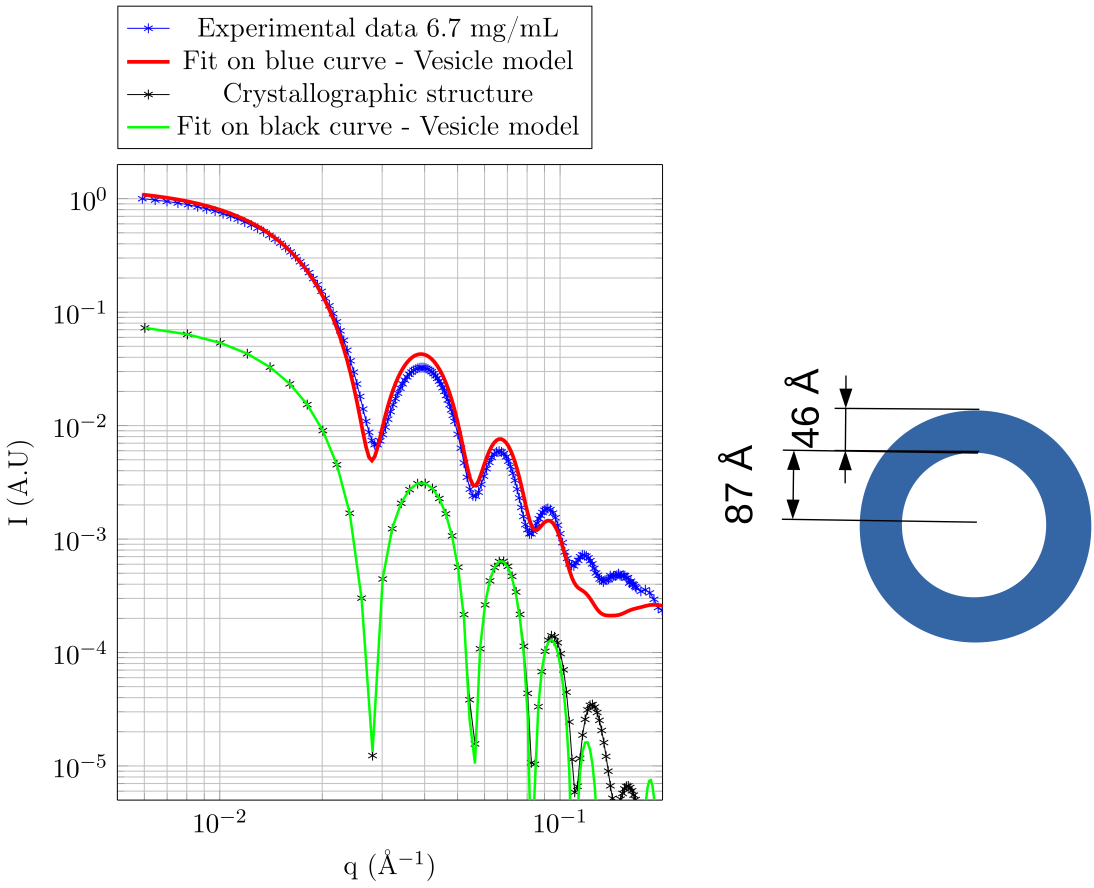


Figure 5-11: Modeling the empty capsid with a vesicle model. *Left*: The fit with a vesicle model is carried out on two sets of data: from the crystallographic structure (black) and from static experiment (blue) at 6.7 mg/mL (pH 5.25, I=0.5 M). The fit on crystallographic data (green) is obtained from a simple vesicle form factor and yields a inner radius of 89 Å and thickness of 44 Å. The fit on experimental data includes a Schulz distribution on the inner radius with constant outer/inner radii ratio. The inner radius is 87 Å, the thickness 46 Å and the polydispersity 0.08. *Right*: The corresponding vesicle with inner radius 87 Å and thickness 46 Å.

The CCMV empty capsid is modeled as a vesicle. The results from a fit on the experimental spectrum of the capsid are shown in fig.5-11. The inner radius is estimated to be 87 Å and the thickness 46 Å. The volume of the capsid can also be calculated as well as the volume of the dimer. Since the capsid volume is 90 times higher than the dimer volume, the latter is found to be equal to V_{dimer}=79 100 Å^3.

5.4.2 Model of the intermediate - Determination of the form factor

5.4.2.1 Description of the object

The intermediate is modeled as a piece of a capsid that results from the projection of an ellipse of semi-axis (a, b) on a sphere of radius R. Fig. 5-12 illustrates how such an object is constructed theoretically. This object will be later assimilated as an "elliptic" cap. If the radius R is fixed (for example equal to the average radius of the CCMV empty capsid), its dimensions only depend upon the value of the semi-axes (a, b) of the ellipse.

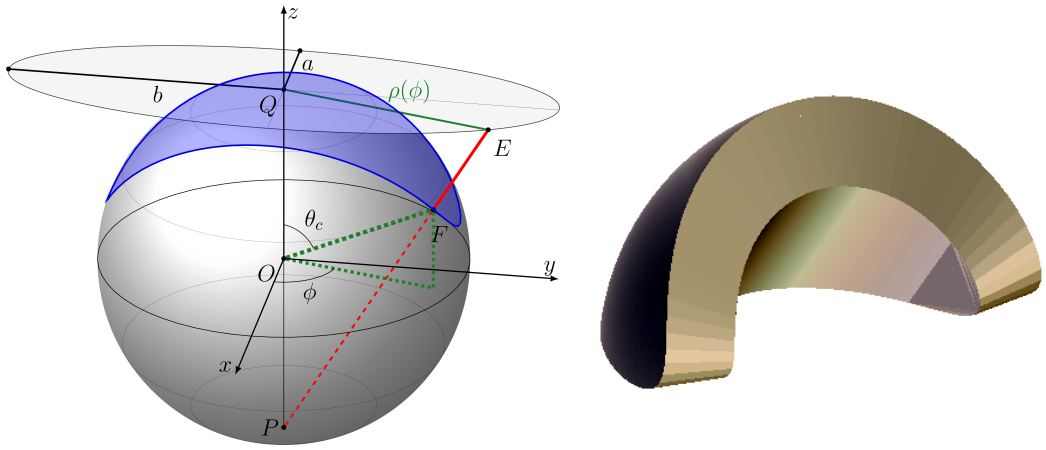


Figure 5-12: Model of the intermediate. *Left*: The intermediate is modeled as a piece of a capsid that results from the projection of an ellipse of semi-axes (a, b) on a vesicle of inner radius R_{min} and outer radius R_{max}. *Right*: Illustration of the resulting elliptic cap. The thickness is 46 Å.

Stereographic projection

Simple geometrical considerations lead to

$$\theta_c(\phi) = 2 \arctan\left(\frac{\rho(\phi)}{2R}\right) \tag{5.3}$$

where $\rho(\phi)$ designates the radial distance in the 2D plane of the ellipse (Fig. 5-12).

For an ellipse of semi-axes a and b ($b > a$):

$$\begin{aligned} \frac{x^2}{a^2} + \frac{y^2}{b^2} &= 1 \\ x = \rho(\phi)\cos\phi, \quad y &= \rho(\phi)\sin\phi \\ \rho(\phi) &= \frac{ab}{\sqrt{b^2\cos^2\phi + a^2\sin^2\phi}} \end{aligned} \tag{5.4}$$

Therefore,

$$\theta_c(\phi) = 2\arctan\left(\frac{ab}{2R}\frac{1}{\sqrt{b^2\cos^2\phi + a^2\sin^2\phi}}\right) \tag{5.5}$$

Volume

The volume of the "elliptic" cap V_{EC} is given by

$$\begin{aligned} V_{EC} &= 4\int_{r=Rmin}^{Rmax}\int_{\phi=0}^{\pi/2}\int_{\theta=0}^{\theta_C(\phi)} r^2\sin\theta \mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi \\ &= \frac{4}{3}(R_{max}^3 - R_{min}^3)\int_{\phi=0}^{\pi/2}\left[1-\cos(\theta_c(\phi))\right]\mathrm{d}\phi \\ &= \frac{4}{3}(R_{max}^3 - R_{min}^3)\left(\frac{\pi}{2} - [F(\phi)]_0^{\pi/2}\right) \end{aligned} \tag{5.6}$$

with

$$F(\phi) = \int_{\varphi=0}^{\phi}\cos(\theta_c(\varphi))\mathrm{d}\varphi \tag{5.7}$$

The function F admits an analytical expression given by

$$F(\phi) = \phi - \frac{2ab}{\sqrt{a^2+4R^2}\sqrt{b^2+4R^2}}\arctan\left(\frac{a\sqrt{b^2+4R^2}\tan\phi}{b\sqrt{a^2+4R^2}}\right) \tag{5.8}$$

Hence,

$$V_{EC} = \frac{4}{3}(R_{max}^3 - R_{min}^3)\frac{\pi ab}{\sqrt{a^2+4R^2}\sqrt{b^2+4R^2}} \tag{5.9}$$

Equation 5.9 establishes a relationship between the volume of the elliptic cap and the semi-axes (a, b) of the ellipse. In practice, R is fixed to the value of the average radius of the CCMV empty capsid. Because the volume of the intermediate is a multiple of the volume of the dimer ($V_{Intermediate}$=V_{EC}=αV_{Dimer}), the object can be described by V_{EC} and one of the semi-axis a or b. This is convenient because in practice, the mass/volume of the intermediate can be fixed to αV_{Dimer} or αM_{Dimer}

and therefore only one of the semi-axis a or b is sufficient to describe the aspect ratio of the object.

5.4.2.2 Spherical harmonics - Fourier transform - Form factor

Real space - Spherical harmonics
Because it is obtained from the projection on a sphere, the object can be described by a series of spherical harmonics:

$$f(r,\theta,\phi) = a(r).\sum_{l=0}^{\infty}\sum_{m=-l}^{l} c_{lm}Y_{lm}(\theta,\phi) = a(r).\sum_{lm} c_{lm}Y_{lm}(\theta,\phi) \tag{5.10}$$

where $Y_{lm}(\theta,\phi)$ represents the spherical harmonics:

$$Y_{lm}(\theta,\phi) = Y_l^m(\theta,\phi) = \sqrt{\frac{(2l+1)}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(cos\theta)e^{im\phi} \tag{5.11}$$

with P_l^m being the associated Legendre polynomials.

$a(r)$ is the gate function

$$a(r) = \begin{cases} 1, & \text{if } R_{min} \leq r \leq R_{max} \\ 0, & \text{otherwise} \end{cases} \tag{5.12}$$

Reciprocal space - Fourier transform - Scattering amplitude
The scattering amplitude is obtained from the Fourier transform of $f(r,\theta,\phi)$:

$$\begin{aligned} A(q,\theta',\phi') = FT\left[f(r,\theta,\phi)\right] &= \int_{\hat{r}} f(r,\theta,\phi)e^{i\hat{q}.\hat{r}}\mathrm{d}\hat{r} \\ &= \int_{\hat{r}} a(r).\sum_{lm} c_{lm}Y_{lm}(\hat{r})e^{i\hat{q}.\hat{r}}\mathrm{d}\hat{r} \end{aligned} \tag{5.13}$$

where the integration is made on the volume of the particle and the hat $\hat{}$ denoting the unit vector.

Using the plane wave expansion,

$$e^{i\hat{q}.\hat{r}} = \sum_{l=0}^{\infty}(2l+1)i^l j_l(qr)P_l(\hat{q}.\hat{r}) \tag{5.14}$$

with j_l being the spherical Bessel functions and P_l the Legendre polynomials. This equation can be rewritten using spherical harmonics with the spherical harmonic

addition theorem,

$$e^{i\hat{q}.\hat{r}} = 4\pi \sum_{l'm'} i^{l'} j_{l'}(qr) Y_{l'm'}(\hat{q}) Y^*_{l'm'}(\hat{r}) \tag{5.15}$$

the superscript * denoting the complex conjugation.

Expressing the scattering amplitude from equation 5.13 with two series of spherical harmonics,

$$\begin{aligned} A(q,\theta',\phi') &= \int_{\hat{r}} \left(a(r).\sum_{lm} c_{lm} Y_{lm}(\hat{r}) \right) \left(4\pi \sum_{l'm'} i^{l'} j_{l'}(qr) Y_{l'm'}(\hat{q}) Y^*_{l'm'}(\hat{r}) \right) \mathrm{d}\hat{r} \\ &= 4\pi \sum_{l'm'} i^{l'} Y_{l'm'}(\hat{q}) \sum_{lm} c_{lm} \int_{\hat{r}} a(r) Y_{lm}(\hat{r}) Y^*_{l'm'}(\hat{r}) j_{l'}(qr) \mathrm{d}\hat{r} \end{aligned} \tag{5.16}$$

The integration can be done by orthogonality of the spherical harmonics, setting $lm{=}l'm'$, so that

$$A(q,\theta',\phi') = 4\pi \sum_{lm} i^l c_{lm} Y_{lm}(\theta',\phi') b_l(q) \tag{5.17}$$

with

$$b_l(q) = \int_0^\infty a(r) r^2 j_l(qr) \mathrm{d}r \tag{5.18}$$

c_{lm} and b_l are coefficients that will be calculated in the next paragraphs.

Scattering intensity

The intensity is expressed as

$$I(q) = \left\langle |A(q,\theta',\phi')|^2 \right\rangle'_\Omega \tag{5.19}$$

with $\Omega' = (\theta',\phi')$ being the solid angle. The square of the norm is obtained from eq. 5.17 .

$$\begin{aligned} \left\langle |A(q,\theta',\phi')|^2 \right\rangle_{\Omega'} &= \left\langle A(q,\theta',\phi') A^*(q,\theta',\phi') \right\rangle_{\Omega'} \\ &= \frac{1}{4\pi} \int_{\Omega'} (4\pi)^2 \left(\sum_{lm} c_{lm} b_l(q) Y_{lm}(\Omega') \right) \left(\sum_{l'm'} c_{l'm'} b_{l'}(q) Y^*_{l'm'}(\Omega') \right) \mathrm{d}\Omega' \\ &= 4\pi \sum_{lm} \sum_{l'm'} c_{lm} b_l(q) c_{l'm'} b_{l'}(q) \int_{\Omega'} Y_{lm}(\Omega') Y^*_{l'm'}(\Omega') \mathrm{d}\Omega' \end{aligned} \tag{5.20}$$

From the orthogonality of the spherical harmonics (when $lm{=}l'm'$), it comes

$$I(q) = 4\pi \sum_{lm} |c_{lm}|^2 |b_l(q)|^2 \tag{5.21}$$

Coefficients $b_l(q)$

$$\begin{aligned} b_l(q) &= \int_0^\infty a(r) r^2 j_l(qr) \mathrm{d}r \\ &= \frac{1}{q^3} \int_0^\infty a(z) j_l(z) z^2 \mathrm{d}z, \ (z = qr) \\ &= \frac{1}{q^3} \int_{z_{min}}^{z_{max}} j_l(z) z^2 \mathrm{d}z, \ (z_{max} = qR_{max}, z_{min} = qR_{min}) \end{aligned} \tag{5.22}$$

The last integral in eq.5.22 admits an analytical expression (using Mathematica) and the coefficients $b_l(q)$ are calculated as follows

$$b_l(q) = \frac{1}{q^3} \left[\frac{\sqrt{\pi}}{2^{l+2}} z^{l+3} \Gamma\left(\frac{l+3}{2}\right) \ {}_1\widetilde{F}_2\left(\frac{l+3}{2}; \frac{2l+3}{2}, \frac{l+5}{2}; \frac{-z^2}{4}\right) \right]_{z_{min}}^{z_{max}} \tag{5.23}$$

where Γ is the Gamma function and ${}_1\widetilde{F}_2$ is a regularized hypergeometric function.

The coefficients b_l can thus be pre-calculated for the q numbers knowing the parameters z_{max} and z_{min}. They do not depend on the volume nor the shape of the intermediate.

Coefficients c_{lm}, c_l

The coefficients c_{lm} result from the projection of the ellipse of semi-axis (a,b) on the sphere of radius R. These coefficients thus contain the information on the shape of the intermediate. They are calculated with the scalar product:

$$\begin{aligned} c_{lm} &= \left\langle Y_{lm}(\Omega), F(\Omega) \right\rangle \\ &= \int_\Omega Y_{lm}^*(\Omega') F(\Omega') \mathrm{d}\Omega' \end{aligned} \tag{5.24}$$

F is the function defined by

$$F(\Omega) = F(\theta, \phi) = \begin{cases} 1, & \text{if } \phi \in [0, 2\pi] \text{and } \theta \in [0, \theta_C(\phi)] \\ 0, & \text{otherwise} \end{cases} \tag{5.25}$$

Thus,

$$c_{lm} = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\theta_C(\phi)} \sqrt{\frac{2l+1}{4\pi}} \sqrt{\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{-im\phi} \sin\theta \mathrm{d}\theta \mathrm{d}\phi \tag{5.26}$$

One can introduce the coefficients c_l defined by

$$c_l = \sum_{m=-l}^{l} |c_{lm}|^2 \tag{5.27}$$

At this stage, two simplifications can be done here :

- from the calculation of the $c_{l(-m)}$,
- from the odd values of the index m.

Calculation of $c_{l(-m)}$: From eq.5.26, and the properties of the associated Legendre polynomials,

$$\begin{aligned} c_{l(-m)} &= \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\theta_C(\phi)} \sqrt{\frac{2l+1}{4\pi}}\sqrt{\frac{(l+m)!}{(l-m)!}}\left\{(-1)^m\frac{(l-m)!}{(l+m)!}P_l^m(\cos\theta)\right\}e^{im\phi}\sin\theta \mathrm{d}\theta\mathrm{d}\phi \\ &= (-1)^m c_{lm}^* \end{aligned} \tag{5.28}$$

Hence,

$$\begin{aligned} c_l &= \sum_{m=0}^{l}|c_{lm}|^2 + \sum_{m=-l}^{0}|(-1)^m c_{lm}^*|^2 - |c_{l0}^2| \\ &= 2\sum_{m=0}^{l}|c_{lm}|^2 - |c_{l0}^2| \end{aligned} \tag{5.29}$$

Odd values of m: Still from eq.5.26, due to the term $e^{-im\phi}$ and because the object is symmetrical, every odd value of the index m leads to c_{lm}=0.

These two points reduce considerably the computation of the coefficients c_{lm} and thus c_l (by a factor of 4).

Form factor

The form factor is the normalized intensity :

$$P(q) = \frac{\sum_l c_l.|b_l(q)|^2}{\sum_l c_l.|b_l(0)|^2} \tag{5.30}$$

From the expression of the intensity in equation 5.21 and $I(q) = V^2P(q)$ where V is the volume of the particle/object, we have

$$V^2 = 4\pi\sum_l c_l.|b_l(0)|^2 \tag{5.31}$$

and an other expression of P(q):

$$P(q) = \frac{4\pi}{V^2}\sum_l c_l|b_l(q)|^2 \tag{5.32}$$

5.4.3 Verification, examples

The implementation of the analytical object was verified in three different ways in order to validate the model.

5.4.3.1 Volume

The first verification concerned the implementation of the object in the algorithm by checking its volume. In the implemented algorithm, the volume of the object was set as a multiple of the volume of the dimer:

$$V_{Intermediate} = V_{EC} = \alpha \times V_{Dimer} \tag{5.33}$$

The volume was also given by the ellipse parameters (a,b) via the coefficients c_l and $b_l(0)$ from equation 5.31.

At the end of the global fitting (see section 5.5), the implementation of the analytical model could readily be checked to be correct from those two expressions of the volume.

5.4.3.2 Form factor: beads and CRYSOL

The second cross-check of the implementation made use of the program CRYSOL [90]. This program uses the multipole expansion (spherical harmonics expansion) of the scattering amplitudes to calculate the scattering pattern of the object determined by its (known) atomic structure. It was found to be appropriate for checking the implementation of the spherical harmonics expansion that was also used in our model. CRYSOL needs the known atomic structure of the object and our object is modeled as a homogeneous elliptic cap. The idea was thus to provide coordinates of beads that represented the electrons inside the elliptic cap. The latter was homogeneous in the sense that the distribution of electron density was homogeneous. The coordinates of the center of the beads were therefore chosen as the nodes of a regular network inside the object.

We tested the implementation of the spherical harmonics expansion on two kinds of objects: a compact and an elongated object that resulted from the projection of a disk and a stretched ellipse respectively as represented in figures 5-13 and 5-14. Both objects had a volume of a pentamer of dimers (α=5).

For both compact and elongated objects, the (X,Y,Z) coordinates were then inserted into a PDB file that served as an input file for CRYSOL. The form factors of the resulting sets of coordinates were then computed by the program and compared with the form factors calculated with our analytical expression.

In figure 5-15, the form factors computed by CRYSOL and by our model are seen to superimpose very well at small angles. The differences occur at higher angles

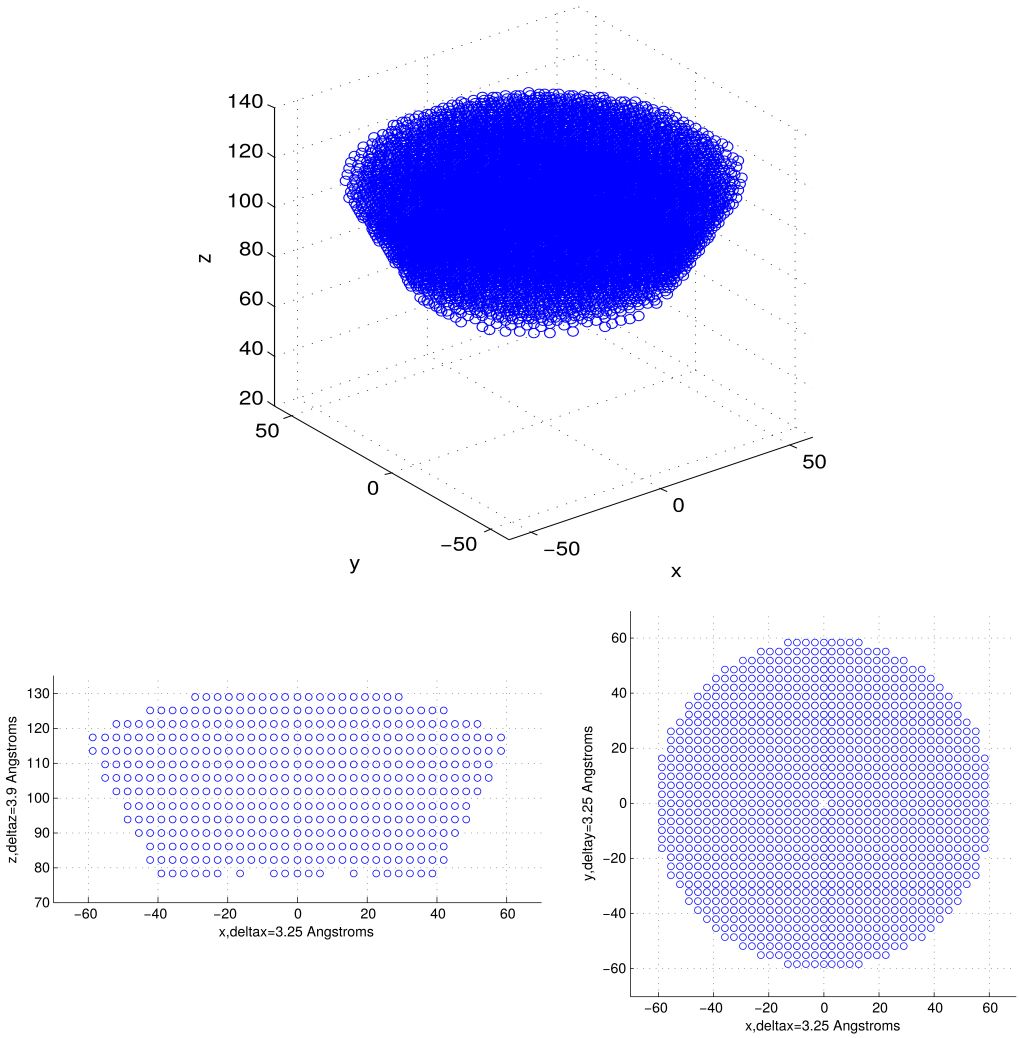


Figure 5-13: Compact object determined by a regularly spaced network of (X,Y,Z) coordinates. The origin is the center of the capsid and the axes are in Å. 9200 beads are displayed. Perspective view of the object is shown in top panel. XY view (bottom left) and XZ view (bottom right) are also represented.

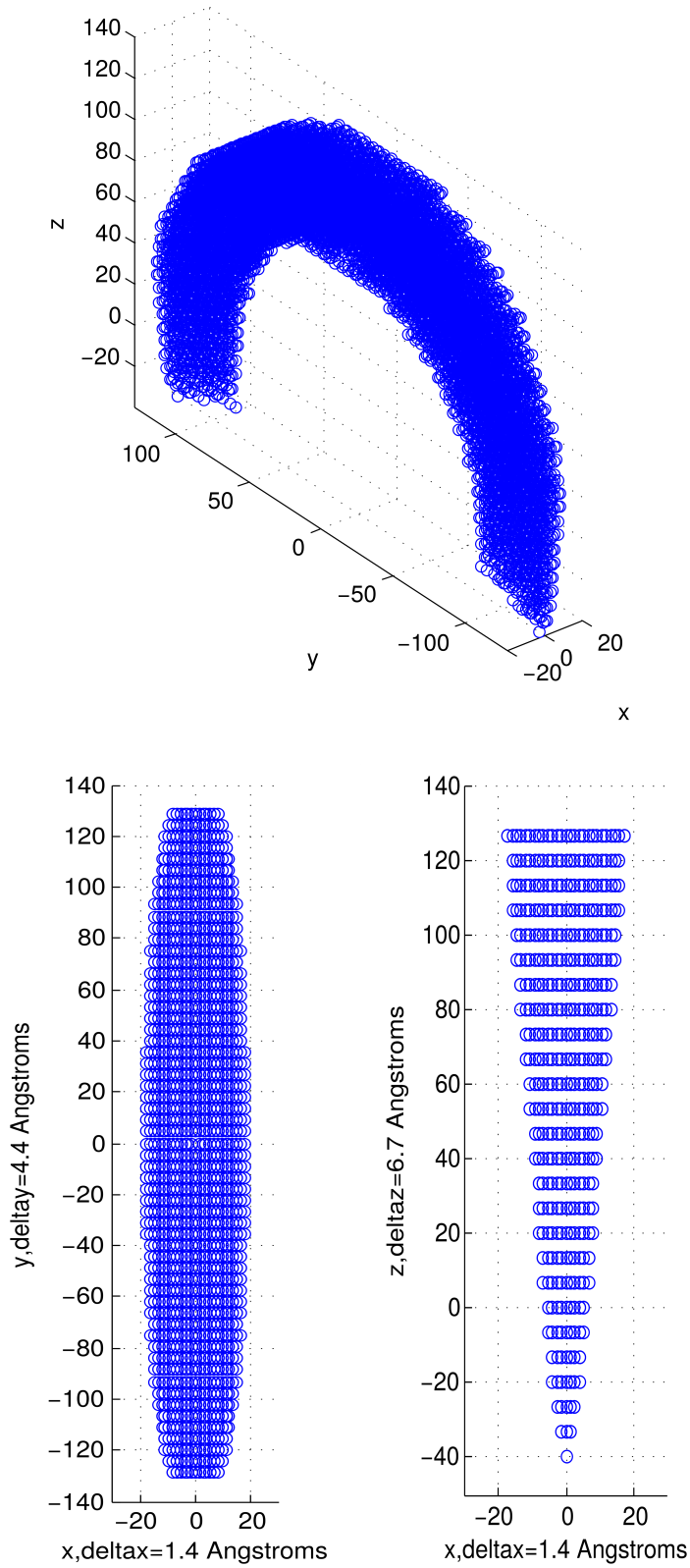
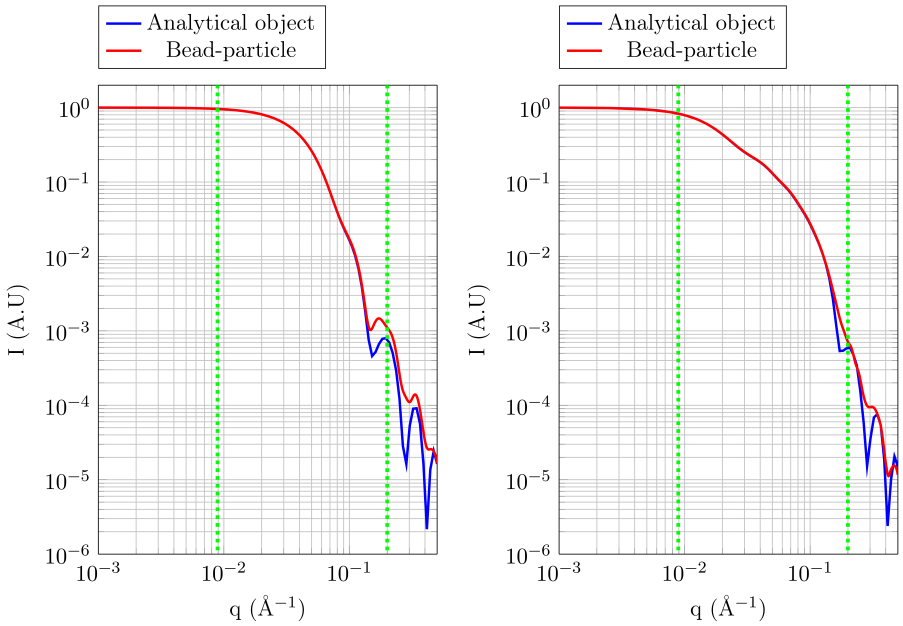


Figure 5-14: Elongated object determined by a regularly spaced network of (X,Y,Z) coordinates. The origin is the center of the capsid and the axes are in Å. 9900 beads are displayed. Perspective view of the object is shown in top panel. XY view (bottom left) and XZ view (bottom right) are also represented.

(>0.15 Å^{-1}). The reasons are still unclear. We were limited in the number of beads for representing the object. It was found that CRYSOL could not compute the form factor of a PDB file that contain more than 10,000 atoms. The distances between two beads in each case were given in figures 5-13 and 5-14 by $\Delta x, \Delta y, \Delta z$. It may not be due to the fact that the distance between two nearest beads is not short enough. Indeed, the distance between particles is roughly $d = \frac{2\pi}{q}$. So, $\frac{2\pi}{\Delta x} = \frac{2\pi}{3.25} \approx 2$ Å^{-1} should be the scattering wavenumber above which the difference between the spectra should be noticeable. Since this difference is for q>0.15 Å^{-1}, the reasons should come from other artifacts, for example to some interpolations used in the CRYSOL program [90, 95].

Nevertheless, the perfect superimposition of the form factors *at small angles* validated the robustness of the implementation of the spherical harmonics expansion in the algorithm.



(a) Compact object (b) Elongated object

Figure 5-15: Comparison between the form factor computed on the analytical objects (blue) and on the corresponding bead-particle (red) for the two objects described in figures 5-13 and 5-14. The form factor of the analytical object is computed using the spherical harmonics expansion described in section 5.4.2.2. The form factor of the corresponding bead-particle is computed by CRYSOL. The form factors superimpose perfectly at small angles which validates the robustness of our model. The dashed vertical lines represent the scattering angle range used for the analyses in the next section.

5.4.3.3 Fit with a crystallographic structure

As an ultimate validation, we performed a comparison based on the geometry of a crystallographic structure and an elliptic cap that fitted this structure in the reciprocal space. In other words, we wanted to know in what extent the fit on a form factor successfully reconstructed the 3D structure of a potential reaction intermediate.

We chose to fit an elliptic cap to the structure of the 30-mer that was presented in Chapter 4. This structure was elongated and consisted of two pentamers of dimers surrounded by C-C dimers. The volume was set to α=15 and the fit was performed on the value of the major semi-axis b of the projected ellipse. As seen in figure 5-16, the model fitted the radius of gyration. The corresponding analytical object is shown in figure 5-17 together with the crystallographic structure.

The results from the fit were satisfying enough to validate the method since the curvature and the aspect ratio of the elliptic cap were respected.

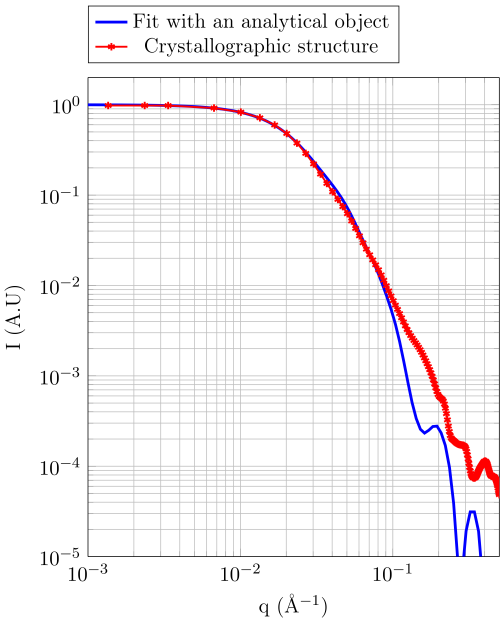


Figure 5-16: Form factor of a crystallographic structure of a piece of the CCMV capsid (elongated 30-mer) fitted with the model of the analytical object. The only fit parameter is the major semi-axis b of the projected ellipse since the volume was set to α=15.

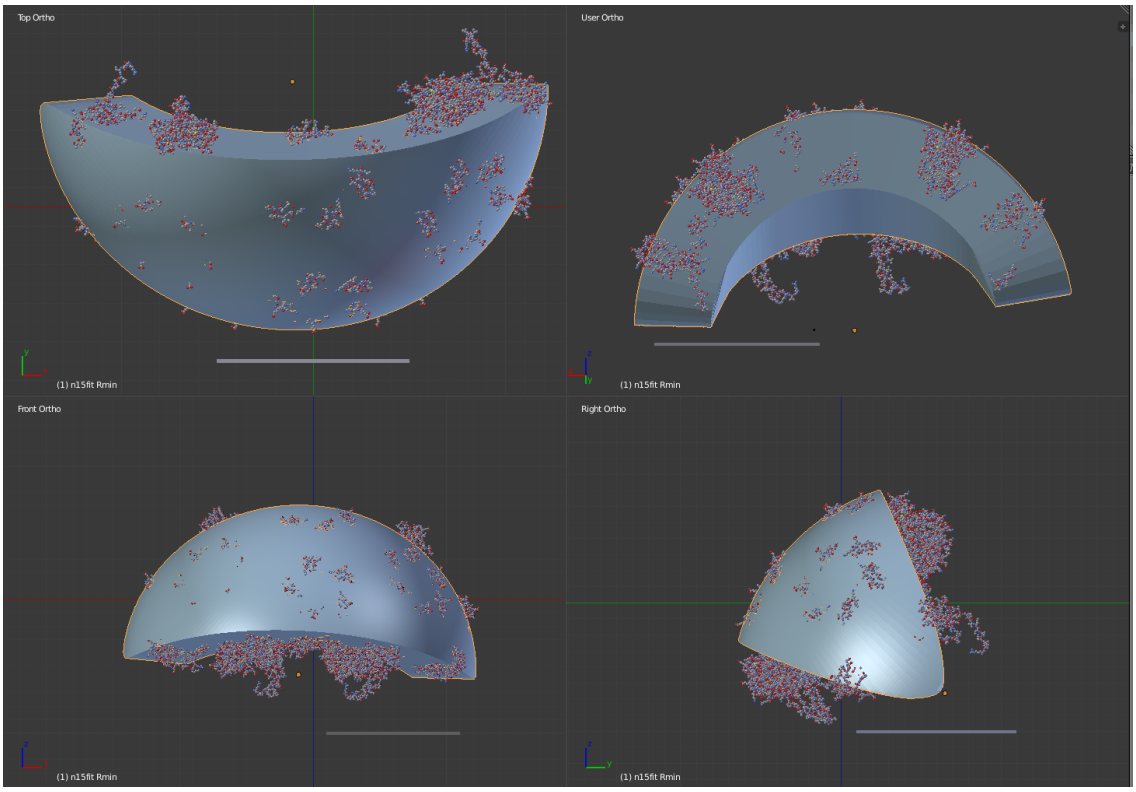


Figure 5-17: Structure of the 30-mer superimposed with the corresponding elliptic object whose form factor fits the spectrum of the 30-mer. All scale bars represent 50 Å.

5.5 Reconstruction of the self-assembly kinetics with a global fitting algorithm

We saw in section 5.3.2 that the SVD analysis suggested that the kinetics of self-assembly could be reconstructed with 3 species :
- the capsid, whose dimensions have been calculated using a vesicle model.
- the dimer, which was found to be the protein subunit from which the capsid is building up.
- and an intermediate, whose mass/volume (via the parameter α) and aspect ratio (via the major semi-axis b) still have to be determined via the model of the elliptic cap.

In this section, the self-assembly process is reconstructed by fitting the data with a kinetic model in order to find the size and shape of the intermediate and the kinetic parameters (stoichiometric coefficients, rate constants $\{k_i^-, k_i^+\}$ and orders of reactions $\{\mu_i\}$).

5.5.1 Global fitting algorithm

The global fitting algorithm was performed on the whole set of data by concatenating the four matrices of concentrations $I^{(k)}(q,t)$ (with k=1,2,3,4 representing the four different concentrations) into a single matrix $I_{cat}(q,t)$. The matrix of intensities was therefore modeled as described in Chapter 3:

$$I_{cat}^{fit}(q,t) = K\mathbf{P'}(q,\{\alpha,b\})\mathbf{c_{cat}}(\{k_i^-,k_i^+,\mu_i\}) \tag{5.34}$$

where $\mathbf{c_{cat}}(\{k_i^-,k_i^+,\mu_i\})$ is the concatenated matrix of concentrations.

$\mathbf{P'}$ comprises the form factors of the capsid, the intermediate and the dimer. The form factor of the capsid was retrieved from the vesicle model mentioned above. The form factor of the dimer was obtained from statics experiment at pH 7.5, 0.5 M NaCl and 0.1 mg/mL. These two form factors were fixed in the model. As for the intermediate, the form factor was computed beforehand, for each value of α tested ($\alpha \in [2;70]$) and several values of b (such that $\max(\theta_C)=9\pi/10$ which corresponds to the greatest aspect ratio that was tested).

The global fitting algorithm used for this study is described as follows:

1. The stoichiometric coefficients to be tested are first fixed in the kinetic model. The parameters to be calculated $\{\xi_i\}$ include the kinetic parameters (forward and backward reaction rates k_i^-, k_i^+ and partial reaction orders μ_i) and the geometrical parameters $\{\alpha, b\}$ of the intermediate to be calculated : $\{\xi_i\} = \{k_i^-, k_i^+, \mu_i, \alpha, b\}$.

2. A loop over all the parameters $\{\xi_i\}$ is implemented using a constrained nonlinear

optimization algorithm already implemented in Matlab (Sequential Quadratic Programming (SQP) and Interior-Point (IP) algorithms were used). The matrix of concentrations $\mathbf{c}(\{k_i^-, k_i^+, \mu_i\})$ is calculated via the resolution of coupled differential equations using the "ode15s" implemented Matlab solver. The matrix $\mathbf{P}'$ is computed as a function of $\{\alpha, b\}$.

3. The residual is computed for several initial sets of parameters $\{\xi_i^0\}$ such that $\mathbf{I} \approx \mathbf{I^{fit}} = K\mathbf{P}'(\{\alpha, b\})\mathbf{c}(\{k_i^-, k_i^+, \mu_i\})$ is optimal.

4. To find the global minimum, the residual is computed for several initial sets of parameters $\{\xi_i^0\}$ that are randomly chosen (uniform distribution) in their chosen range defined with lower and upper bounds, $\{\xi_i^{min}, \xi_i^{max}\}$ within the first loop. Typically 30-50 initial points were found to be sufficient to find the global minimum.

5. Once the residual χ^2 is found to be a global minimum, its value is temporarily stored and steps 2 to 4 are repeated for other sets of stoichiometric coefficients.

6. The values of the stoichiometric coefficients and the residual χ^2 are stored along with the matrices $\mathbf{c}$ and $\mathbf{P}'$ and the set of parameters $\{\xi_i\}$.

5.5.2 Results from the global fitting

The complexity of such a study was that the intermediate (from the parameters $\{\alpha, b\}$) as well as the kinetic reactions (parameters $\{k_i^-, k_i^+, \mu_i\}$) were both undetermined. In particular, the parameters $\{k_i^-, k_i^+, \mu_i\}$ depend upon the kinetics model chosen to fit the data. One had to devise a kinetic model that was coherent with the presence of the three species mentioned earlier.

We therefore tested two different kinetics model: the nucleation-elongation model as proposed by Zlotnick *et al.* [24, 37, 96] and a more straightforward three-state model that was used by Tresset *et al.* for testing the cooperativity in the assembly of a norovirus [27].

The fit was assessed quantitatively from the residual χ^2 and the R_{factor}, a figure of merit used in crystallography [97] defined as

$$R_{factor} = \sum_{m,n=1}^{M,N} |I_{mn} - I_{mn}^{fit}| \left(\sum_{m,n=1}^{M,N} |I_{mn}| \right)^{-1} \tag{5.35}$$

5.5.2.1 Nucleation-elongation model

Description

The nucleation-elongation mechanism (see fig. 5-18) is first described by a first reaction (nucleation) in which α=n protein subunits (here, dimers D) build up a nucleus

(intermediate I). This reaction is supposed to be reversible with a forward reaction rate k_1^+ and a backward reaction rate k_1^-. The second step of the reaction is the succession of several reactions where, in each reaction, a dimer is added after one another (elongation) until the capsid is formed. The second step is assumed to be fast. This is why in our model each of the successive reactions are modeled with the same forward reaction rate k_2^+ and a backward reaction rate k_2^-. The corresponding

$$
\begin{aligned}
\mathrm{nD} &\underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \mathrm{I_n} \\
\mathrm{I_n} + \mathrm{D} &\underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{I_{n+1}} \\
\mathrm{I_{n+1}} + \mathrm{D} &\underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{I_{n+2}} \\
&\vdots \\
\mathrm{I_{89}} + \mathrm{D} &\underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{C}
\end{aligned}
$$

Figure 5-18: Nucleation-elongation reaction scheme. Note that we used an other notation α=n.

set of coupled equation is given in equation 5.36.

$$
\begin{aligned}
\frac{dc_2}{dt} &= k_1^+ c_1^{\mu_n} - (k_1^- + k_2^+ c_1)c_2 + k_2^- c_3 \\
\frac{dc_i}{dt} &= k_2^+ c_1 c_{i-1} - (k_2^- + k_2^+ c_1)c_i + k_2^- c_{i+1}, \quad for \;\; i = 3...m+1 \\
\frac{dc_{m+2}}{dt} &= k_2^+ c_1 c_{m+1} - k_2^- c_{m+2} \\
\frac{dc_1}{dt} &= -n\frac{dc_2}{dt} - \sum_{j=3}^{m+2} \frac{dc_j}{dt}(n+j-2)
\end{aligned}
\tag{5.36}
$$

where the index i=1 corresponds to the dimer, i=2 the intermediate composed of n dimers and i=m+2 the capsid. The indices $i \in [3; m+1]$ are for the intermediates composed of $n+1$ to $m+n$-1=89 dimers.

To be coherent with the model, these intermediates were assumed not to contribute during the kinetics. Only the n-dimer intermediate was assumed to have a lifetime long enough to be detected. This was simply done by removing their contribution in the matrix of concentrations **c**.

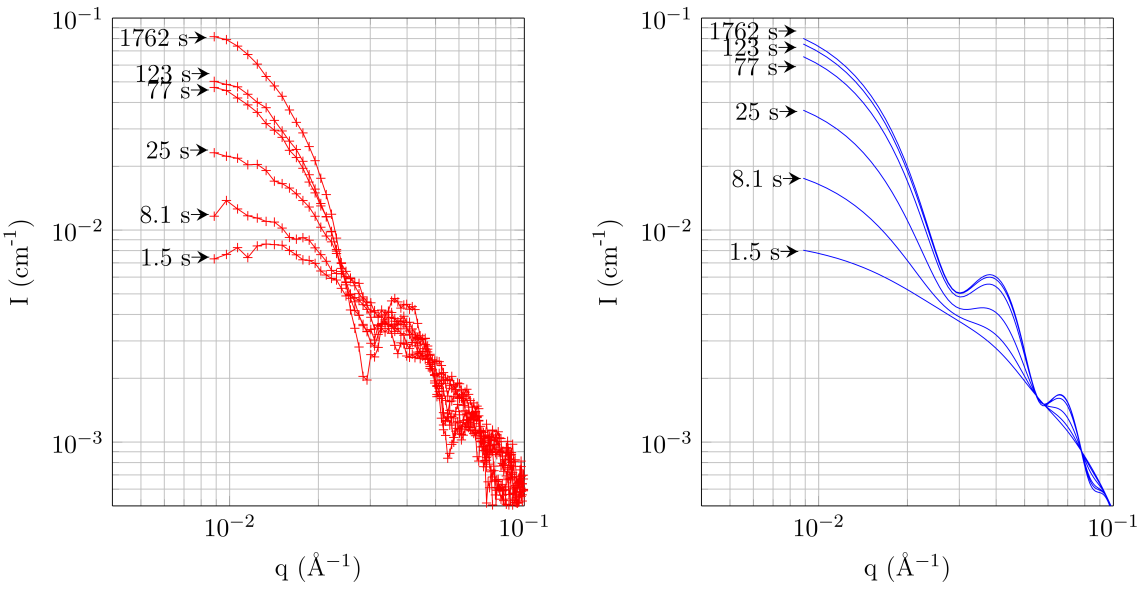
Nucleation-elongation scheme: results

The results from the global fitting are shown in table 5.1. The value of the residual

χ^2 is minimum for n=66, meaning that in the nucleation-elongation scheme, the most potential intermediate is composed of 66 dimers (the capsid is composed of 90 dimers). This is far from the value of n=5 established by Zlotnick *et al.* [24]. We emphasize on the fact that the upper and lower bounds of the undetermined parameters (see step 4 of the algorithm) were kept large enough to ensure that the model would fit consequently. Figure 5-19 also qualitatively shows that the nucleation-elongation scheme is not satisfying for reproducing well the oscillations from the form factor of the capsids.

n	m	χ^2	Rfactor(%)
2	88	28.75	29.9
5	85	23.51	61.7
10	80	20.08	56.0
18	72	16.30	49.5
24	66	13.68	44.2
30	60	11.41	40.5
36	54	8.79	32.9
40	50	7.34	30.9
44	46	6.02	27.1
48	42	5.42	24.5
54	36	3.81	19.9
60	30	3.20	20.1
66	24	2.23	14.9
70	20	3.11	20.3

Table 5.1: Nucleation-elongation mechanism. Results from global fitting for different combinations $\{n, m\}$ with n=α being the size of the nucleus (number of dimers) and $n + m$=90.



(a) Experimental data

(b) Nucleation-elongation model

Figure 5-19: Nucleation-elongation model. (a) Experimental data at 0.48 mg/ml. (b) Corresponding model that fit the data with n=66 (χ^2=2.23).

5.5.2.2 Three-state model

Description

We then looked over a three-state mechanism, which was simply based on a two-step reaction involving the three species (see scheme 5-20). With such a scheme, because a capsid is composed of 90 dimers, the constraint $\alpha\beta + \gamma = 90$ has to be respected.

$$\alpha \mathrm{D} \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \mathrm{I}$$

$$\beta \mathrm{I} + \gamma \mathrm{D} \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{C}$$

Figure 5-20: Three-state reaction scheme. The stoichiometric coefficients are linked via the relation $\alpha\beta + \gamma = 90$.

This scheme is probably the most simple reaction that one can consider with the three species involved. It has the advantage to also test the cooperativity of the mechanism. By "cooperativity", it is meant that at least two intermediates ($\beta > 1$) interact with each other and complete the capsid likely with the addition of dimers.

The corresponding set of coupled equations is given below (eq 5.37).

$$\begin{aligned}\frac{dc_1}{dt} =&\alpha(k_1^- c_2 - k_1^+ c_1^{\mu_\alpha}) + \gamma(k_2^- c_3 - k_2^+ c_1^{\mu_\gamma} c_2^{\mu_\beta}) \\ \frac{dc_2}{dt} =&\beta(k_2^- c_3 - k_2^+ c_2^{\mu_\beta} c_1^{\mu_\gamma}) + k_1^+ c_1^{\mu_\alpha} - k_1^- c_2 \\ \frac{dc_3}{dt} =&k_2^+ c_1^{\mu_\gamma} c_2^{\mu_\beta} - k_2^- c_3\end{aligned} \tag{5.37}$$

where the index i=1 corresponds to the dimer, i=2 the intermediate composed of α dimers and i=3 the capsid.

In the global fitting algorithm, the values of a set $\{\alpha, \beta, \gamma\}$ were first fixed. The partial order of reaction μ_β was set accordingly to β in order to be coherent with the cooperativity of the second step of the overall reaction.
The partial order of reaction μ_γ was however set to 1. Indeed, for numerical considerations, it was noticed that a value of μ_γ >10 prevents the algorithm to converge. A value of μ_γ=1 for the second step of the reaction means that the dimers do not contribute much at this stage of the kinetics, which is a plausible hypothesis. It suggests that this stage of the reaction is indeed dominated by the concentration of the intermediate.
μ_α is in fact the reaction order of the first step of the overall assembly reaction. Since a n-body reaction occurring in a single step (with n more than 3) is rare (the model of Zlotnick *et al.* actually assumed a 5-body reaction occurring in a single step), the partial order of reaction μ_α was set to belong to the interval $[1; 3]$.

Three-state scheme: results
It is important to note that we tested several combinations of $\{\alpha, \beta, \gamma\}$ that took into account several possibilities of building up the capsid with pieces of capsid portions represented by the intermediates. We particularly distinguished the case $\beta > 1$ (the second step of the reaction is cooperative) and $\beta = 1$.

As shown in table 5.2, the case $\beta > 1$ did not yield a better fit than the nucleation-elongation mechanism as it was previously seen. However, the case $\beta = 1$ (table 5-21) better fitted the data for $\alpha \in [35; 48]$.

These results suggest that the second step of the reaction is not likely a cooperative mechanism. Two half-capsid intermediates do not assemble together to form the full capsid. Instead, one intermediate of a half-capsid size, composed of 35-48 dimers, is likely to complete into a capsid with the addition of dimers. Furthermore, the addition of dimers does not occur one after the other with the same frequency until the capsid is formed, as it is assumed in an elongation process. It is rather presumed that the capsid completion becomes onerous in terms of energy as the size of the particle is increasing. This should be due to the fact that, as the size is increasing, the energy and entropy needed for the additional dimers to bind with the correct

α	β	γ	χ^2	Rfactor (%)
2	24	42	12.69	39.8
2	45	0	12.85	38.2
5	12	30	10.85	37.7
10	3	60	4.72	24.6
10	6	30	7.56	30.6
18	3	36	4.14	21.9
18	4	18	5.50	24.7
24	3	18	3.90	20.0
30	2	30	2.47	16.9
30	3	0	3.81	19.5
36	2	18	2.33	14.8
38	2	14	2.98	16.2
40	2	10	2.21	14.5
44	2	2	2.24	14.1

Table 5.2: Three-state cooperative model. Results from global fitting for different combinations $\{\alpha, \beta, \gamma\}$ with $\beta > 1$.

orientation is cost expensive.

It may also be possible that the completion into a capsid is induced by the addition of some oligomers that are smaller than the half-capsid intermediate, so that they cannot be detected and considered in the reaction.

Table 5.3 specifies the kinetic parameters obtained from the fit for different values of α for the case $\beta = 1$. The calculated value of μ_α suggests that the first reaction of the three-state mechanism is a first-order reaction. It is worth noticing the values of the backward rate constants k_1^- and k_2^-. The obtained values are indeed high, which suggests that the reactions are strongly reversible.

Figures obtained from the fit

To further describe and illustrate the results from the global fitting, we plot the modeled intensities and the form factors of the three species involved in the kinetics in figure 5-22.

The modeled patterns from the fit are also superimposed with the experimental patterns in figure 5-23 for the value α=37. It can be seen that the modeled spectra superimpose well the kinetics from about 25 s. The difference from 0 s to 25 s (7 out of 54 spectra for each kinetics) should be due to the fact that smaller oligomers are formed during this phase. These smaller oligomers cannot be reconstructed because of the lack of information at early times.

Finally, the mass fractions of the three species established from the modeled concentrations from the fit are plotted in figure 5-24.

α	β	γ	χ^2	Rfactor(%)
5	1	85	2.45	15.1
10	1	80	2.27	13.6
18	1	72	2.13	13.8
24	1	66	1.86	12.4
30	1	60	1.71	11.8
36	1	54	1.69	11.7
37	1	53	1.67	11.6
38	1	52	1.66	11.7
40	1	50	1.67	11.9
44	1	46	1.69	12.0
48	1	42	1.68	11.8
52	1	38	1.69	11.9
56	1	34	1.72	12.0
60	1	30	1.77	12.3
66	1	24	1.87	12.4
70	1	20	2.00	13.1

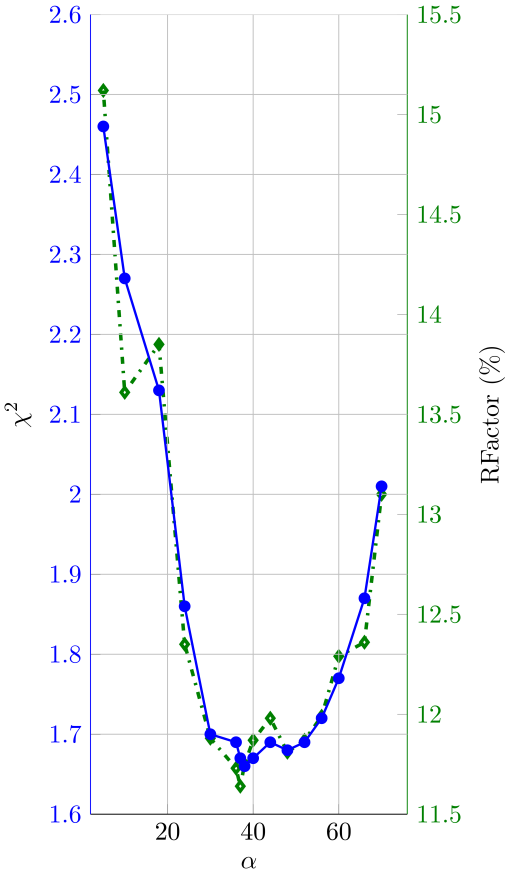


Figure 5-21: Three-state model. *Left* : Results from global fitting for different values of α and for $\beta = 1$. *Right* : Corresponding plot showing the two figures of merit as a function of α. The residual is minimum for $\alpha \in [37; 48]$.

α	β	γ	k_1^+	k_1^-	k_2^+	k_2^-	R_g(Å)	μ_α	μ_β	μ_γ
5	1	85	9.8e-3	4.0e-1	1.1e-4	8.5e-3	92	1.193	1	1
10	1	80	4.1e-3	3.8e-1	2.9e-4	7.5e-3	60	1.024	1	1
18	1	72	5.1e-4	1.7e-1	6.2e-4	7.6e-3	69	1.173	1	1
24	1	66	1.5e-3	3.3e-1	4.9e-4	6.7e-3	83	1.076	1	1
30	1	60	7.7e-4	2.1e-1	7.2e-4	8.1e-3	87	1.043	1	1
37	1	53	9.9e-4	3.4e-1	7.2e-4	7.0e-3	94	1.059	1	1
40	1	50	4.6e-4	1.4e-1	7.2e-4	7.5e-3	96	1.027	1	1
48	1	42	5.4e-4	1.7e-1	7.3e-4	7.1e-3	101	1.000	1	1
60	1	30	1.6e-4	1.2e-1	7.1e-4	5.4e-3	108	1.000	1	1
70	1	20	1.3e-4	1.2e-1	7.4e-4	5.0e-3	112	1.216	1	1

Table 5.3: Three-state model. Calculated parameters from global fitting for the associated values of χ^2 and R_{Factor} in table 5-21. Rate constants are k_1^+ $(\mu M^{1-\mu_\alpha}.s^{-1})$, k_1^- (s^{-1}), k_2^+ $(\mu M^{1-\mu_\beta-\mu_\gamma}.s^{-1})$ and k_2^- (s^{-1}).

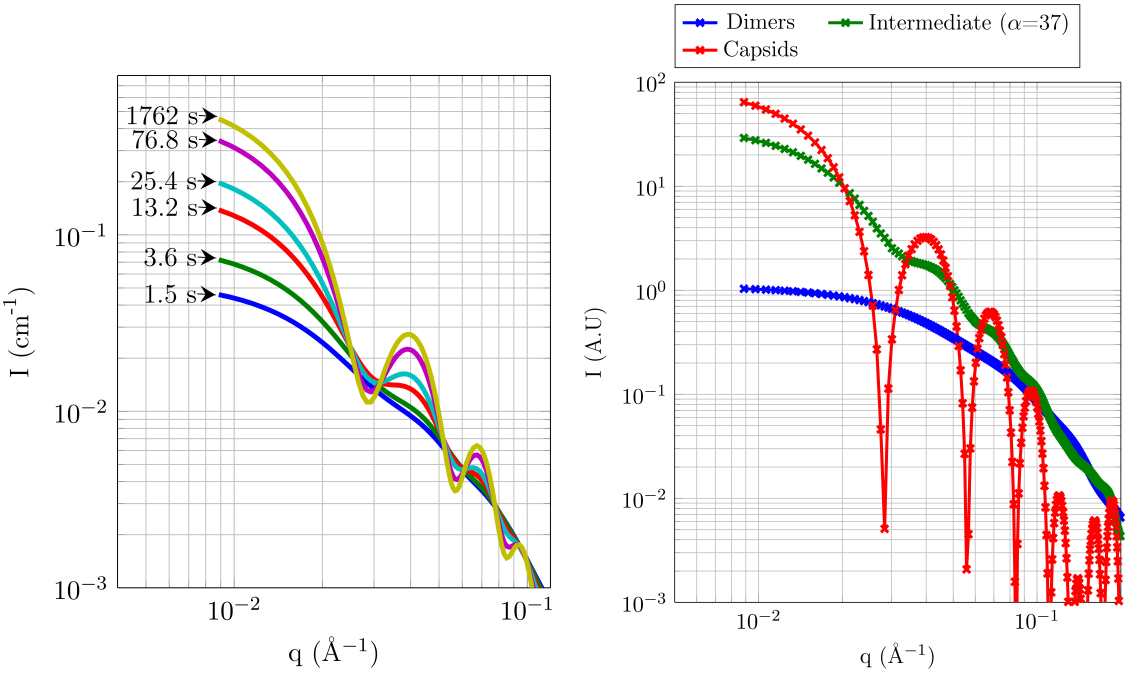
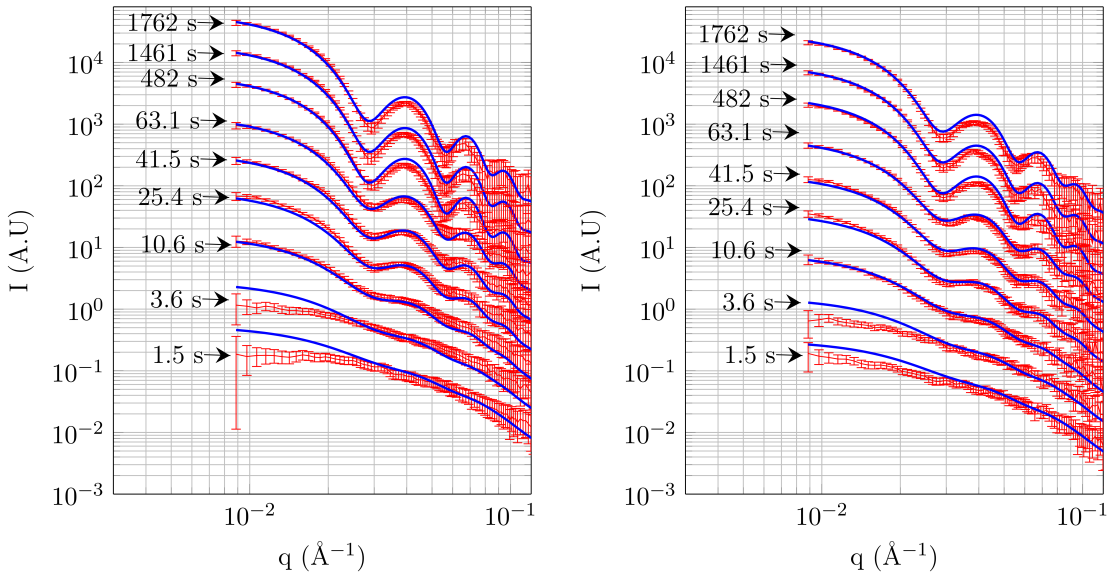
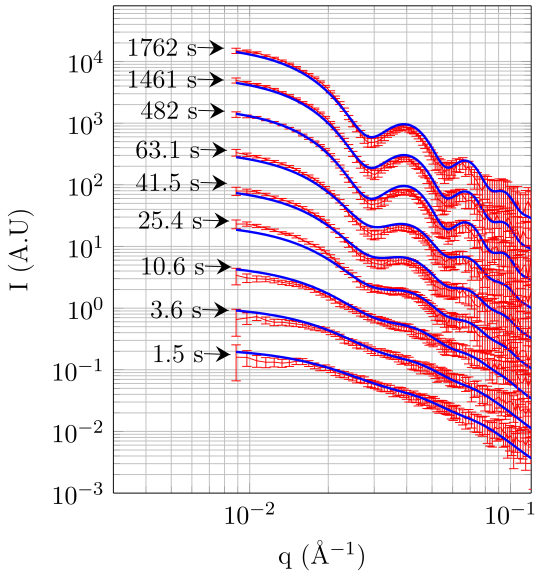


Figure 5-22: Three-state model results with α=37, β=1 (χ^2=1.67). *Left*: The modeled intensities plotted at different times of the kinetics. *Right*: Form factors of the three species weighted by the numbers of dimers.

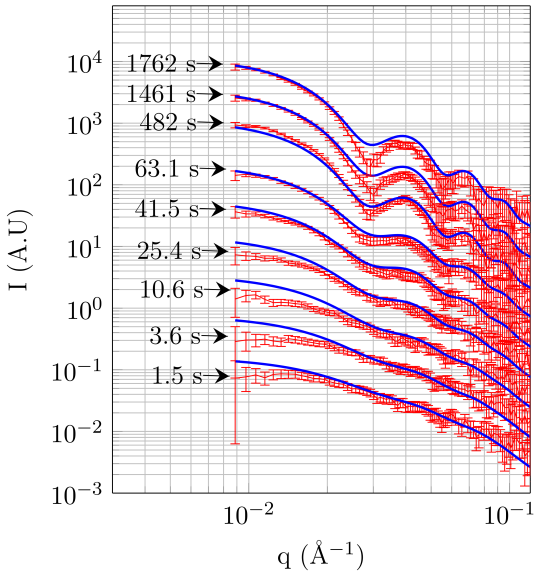


(a) 1.49 mg/mL

(b) 0.96 mg/mL

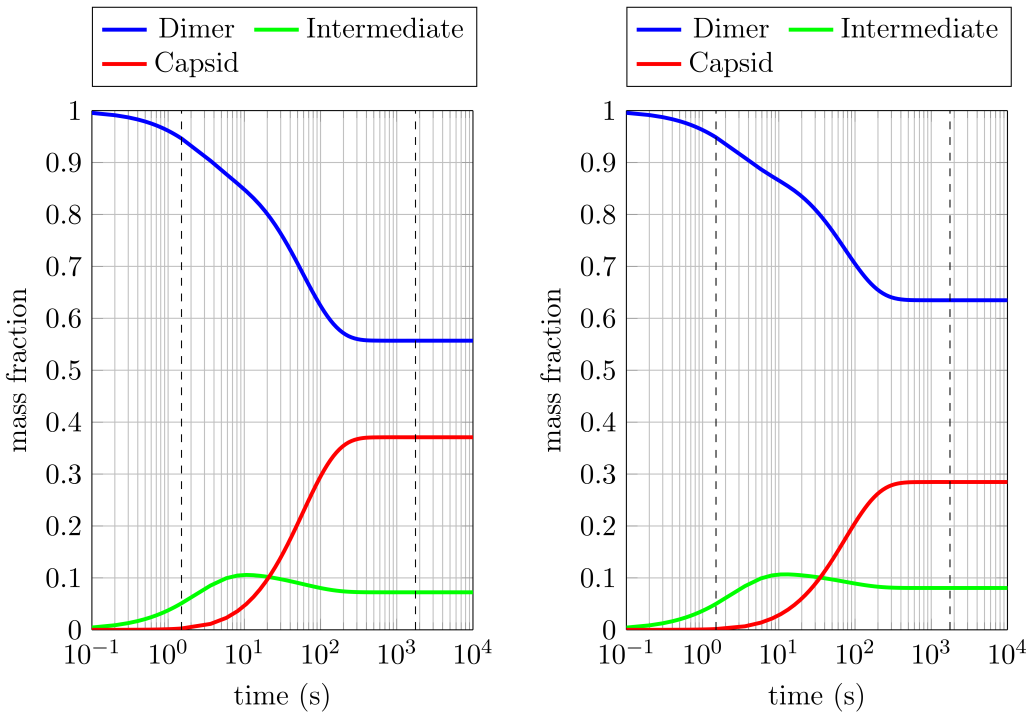


(c) 0.67 mg/mL



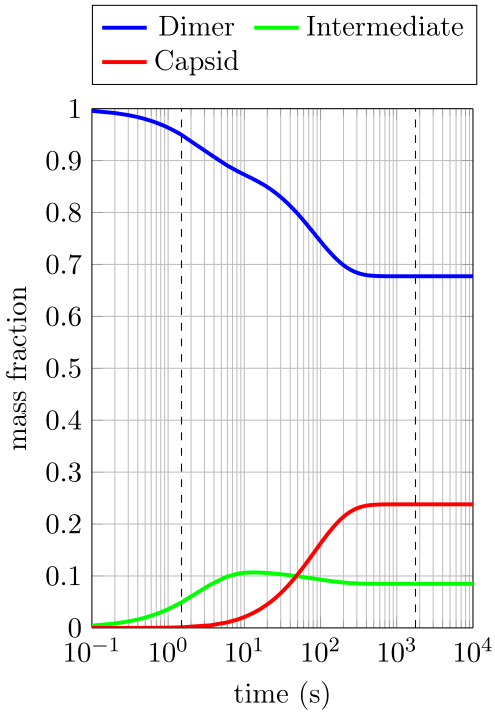
(d) 0.48 mg/mL

Figure 5-23: Three-state model. TR-SAXS patterns fitted with the model with α=37, β=1 (χ^2=1.67) for the four concentrations. The patterns are shifted for clarity.

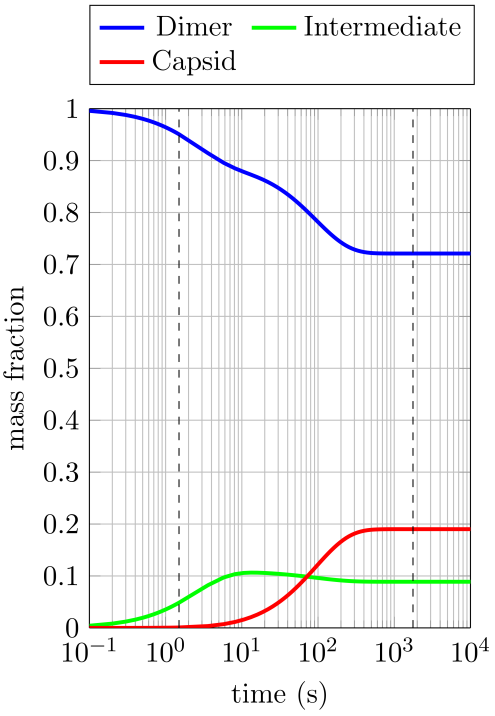


(a) 1.49 mg/mL

(b) 0.96 mg/mL



(c) 0.67 mg/mL



(d) 0.48 mg/mL

Figure 5-24: Three-state model. Mass fractions obtained from the fit with α=37, β=1 (χ^2=1.67) for the four concentrations. The dashed vertical lines represent the time frame of the experiment.

5.5.3 Structure of a potential intermediate

The above results suggest that the intermediate is of a the size of a half-capsid, on average. From the values of the residual χ^2 and the R_{factor}, the mass/volume of the average intermediate is about 35-48 dimers. The aspect ratio of the projected ellipse is also given from the results of the fit, which is found to be close to $(a{:}b)$=(1:1).

From these information, I proposed a structure of such a potential intermediate established from the crystallographic structure of the CCMV capsid, which is illustrated in figure 5-25. Such an intermediate is compact and is composed of 37 dimers. These dimers are organized in such a way that only pentamers of dimers (PODs) and hexamer of dimers (HODs) are completely formed. Every dimer attached to this structure are completing both PODs and HODs. I chose this structure because the CCMV capsid is exclusively made up of 12 pentameric and 20 hexameric sites. These capsid domains were also found to be the most mechanically stable quasi-rigid units from molecular dynamics simulations on the CCMV capsid [98].

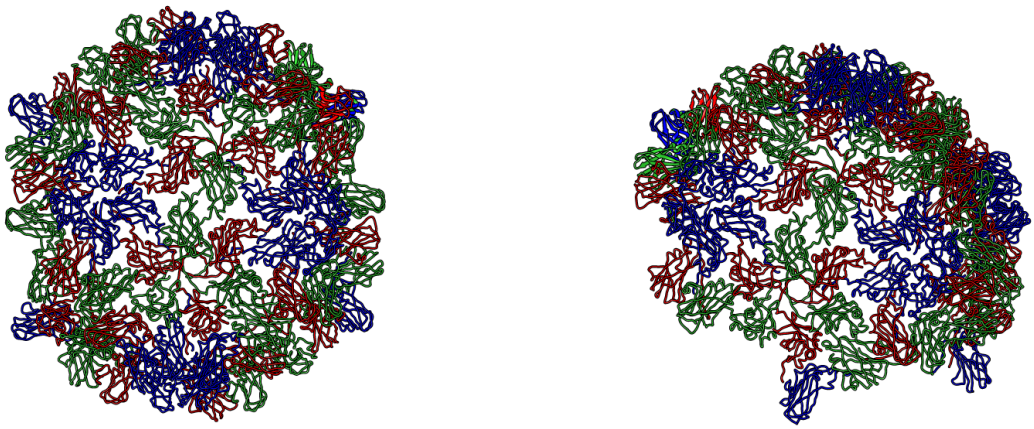


Figure 5-25: Two different views of the structure of a potential intermediate established from the crystal structure of the CCMV capsid (PDB 1ZA7). This structure is composed of α=37 dimers and exhibits 4 pentameric and 6 hexameric sites.

Figure 5-26 presents the form factor of the selected structure computed by CRYSOL superimposed with the form factors of the elliptic caps that best fitted the TR-SAXS data. It shows that the form factor fits within a certain envelope determined by the whole set of elliptic caps form factors, up to 7.10^{-2} Å^{-1}.

Figure 5-27 represents the crystallographic structure of the proposed 37-dimers intermediate from a PDB file together with the corresponding 37-dimers analytical object whose spectrum fitted the data well (χ^2=1.67). It shows that the aspect ratio and the curvature are respected.

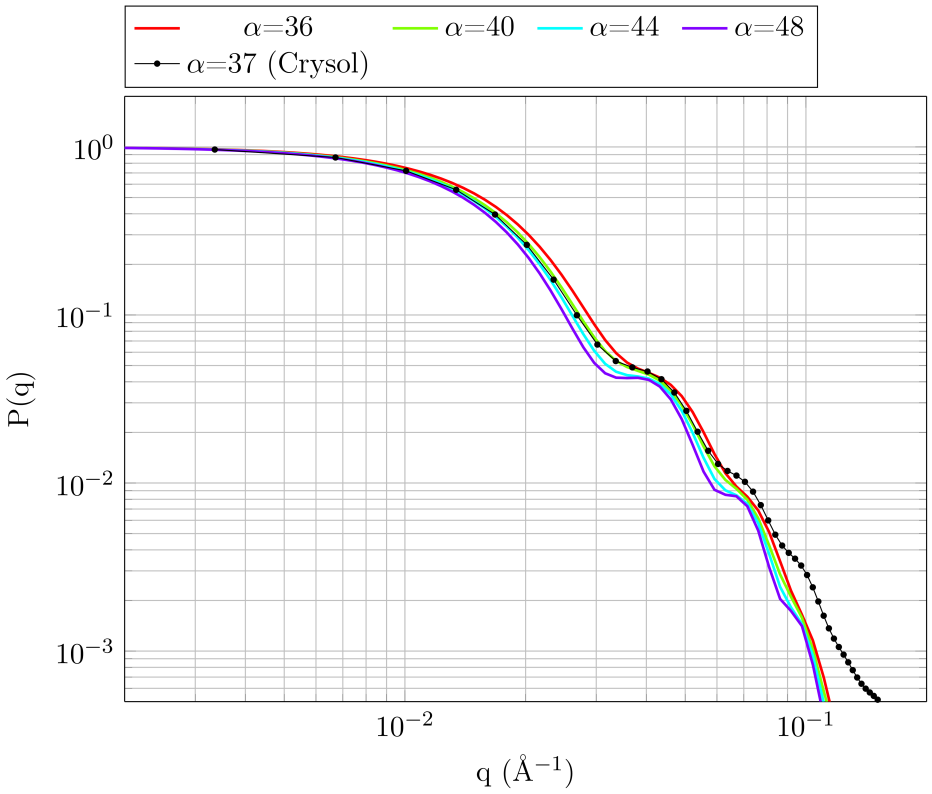


Figure 5-26: Form factor of the crystallographic model (α=37) presented above calculated with CRYSOL [90] superimposed with the form factors of the analytical objects that best fit the data (see table 5-21). The radii of gyration were (for increasing order of α) 92, 94, 96, 99 and 101 Å.

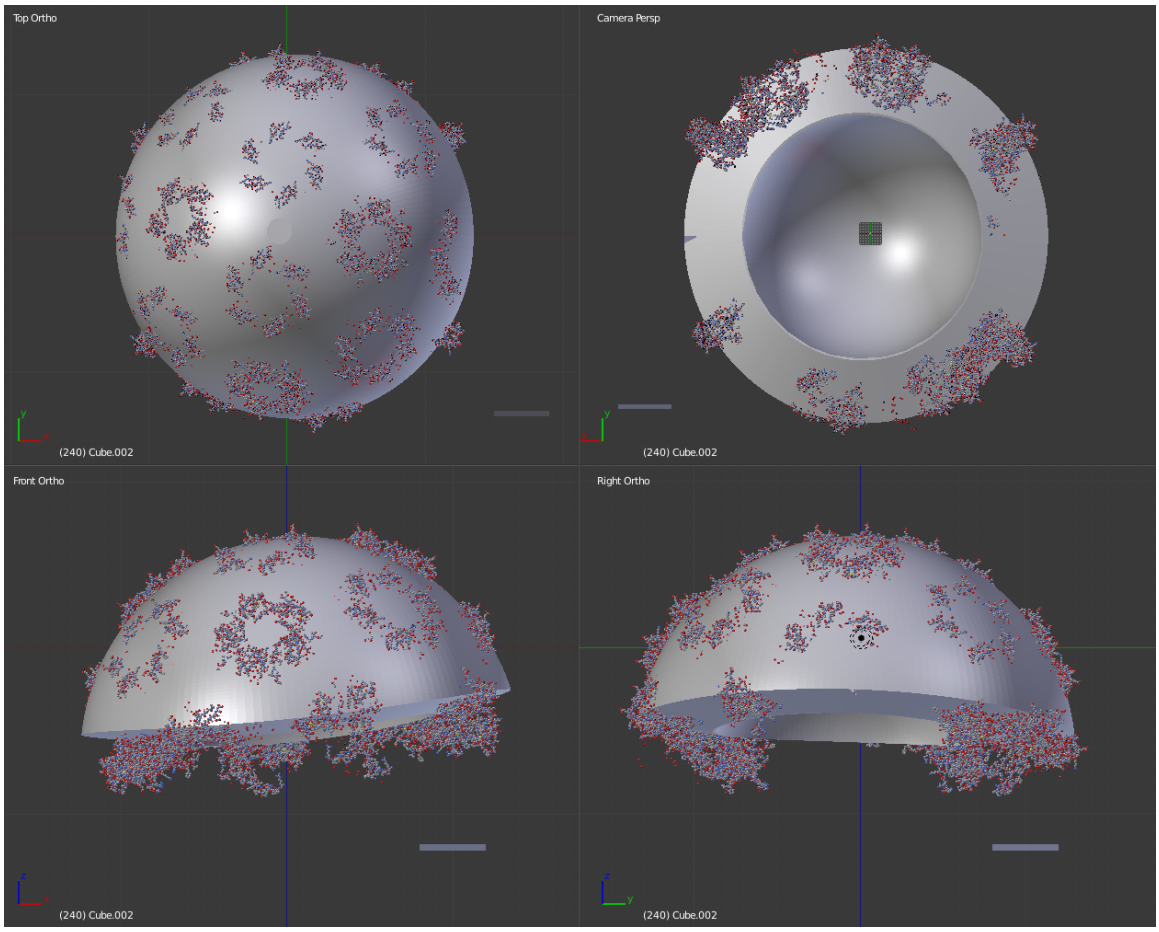


Figure 5-27: Different views of the object obtained from the fit for α=37 superimposed with the crystallographic structure shown in 5-25. All scale bars are 40 Å. The image was made with the Blender software.

5.5.4 Comparison self-assembly/disassembly mechanisms

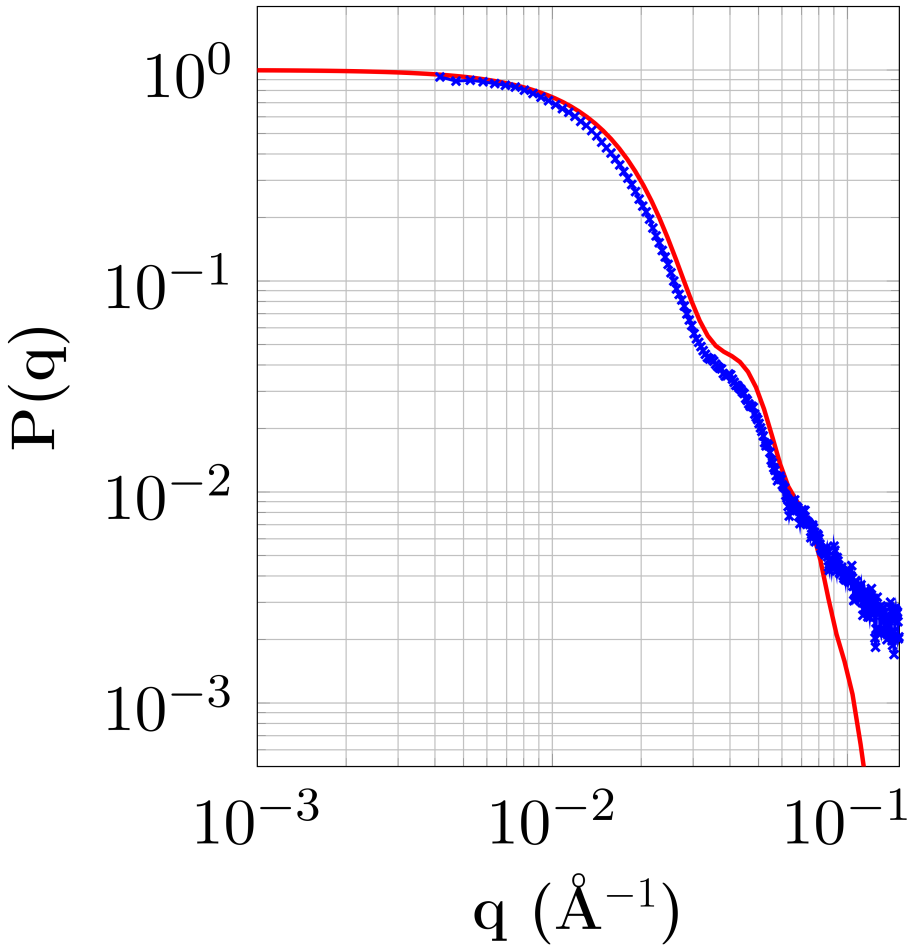


Figure 5-28: Comparison of the two half-capsid sized intermediates detected in the self-assembly (red curve) and disassembly (blue curve) kinetics. The red curve is obtained with the elliptic cap model (R_G=94 Å). The blue curve is computed from least squares fitting algorithms (R_G=104 Å).

We recently reported [89] the detection of an half-capsid size intermediate in the disassembly kinetics of CCMV VLPs of similar size of the intermediate that we detected in the self-assembly process. Indeed, the average mass was estimated from the global fitting algorithms and was about of 30-40 dimers for the big intermediate detected in the disassembly process and of about 35-50 dimers for the one detected in the self-assembly process. Whether the two half-capsid intermediates had the exact same mass and structure was still uncertain. In Fig. 5-28, we plotted the two form factors of the corresponding intermediates that we found to be close. The small difference could be interpreted by the fact that the objects were in different conditions of pH. Indeed, in the self-assembly experiment, the pH was set to 5.25 whereas it was set to 7.5 in the disassembly one. We also put forward that in pH 7.5, the de-

protonation of acidic residues enhanced the electrostatic repulsion between proteins subunits as it was reported in previous investigations [16, 88]. This may explain why the radius of gyration is larger in the case of the disassembly (R_G=104 Å) than in the self-assembly kinetics (R_G=94 Å).

Both the similar average mass and the resemblance from the structure however support our findings in the half-capsid sized intermediate detected in both self-assembly and disassembly process.

5.5.5 Discussion

The main advantages of the mathematical model to represent the intermediate were that some additional constraints could be added in the algorithm. The volume and the mass of the object were indeed fixed parameters and only one parameter was allowed to vary. This parameter was the major semi-axis of the projected ellipse and determined the aspect ratio of the elliptical cap. At the scattering vector range q of the TR-SAXS experiments (small angles), the difference between the form factor of the elliptical cap and the real form factor of the intermediate was not likely to be high, especially for large objects. Furthermore, the algorithm was better fitting at the range of smaller angles where the signal-to-noise ratio was higher, since the residual was the χ^2 that was somehow linked to this ratio. The range of smaller angles was actually the Guinier range which contained the information on the mass/volume of the object and the overall shape (also given by the radius of gyration). The elliptical cap model was therefore suitable for this study because the model only had to provide the best fit by determining the aspect ratio given that the mass/volume was embedded in the model accordingly.

The price to pay to study complex assemblies involving unknown reaction orders was to study the reaction at different initial concentrations. The different kinetics data sets had then to be embedded into a single data set to constrain the model and to determine the reaction orders. Four kinetics data sets at different concentrations were collected : 0.48 , 0.67, 0.96 and 1.49 mg/mL. We found out that a model integrating these four concentrations and involving a first-order reversible reaction followed by a reversible reaction of overall order of reaction 2 fitted the data well.

Zlotnick *et al.*, when they probed the assembly of CCMV empty capsids with combined size exclusion chromatography and light scattering, observed that pentamers of dimers (PODs) were produced during the reaction. We draw attention that the experimental conditions between the studies from Zlotnick *et al.* and ours essentially differed in the buffers used in the study. Zlotnick *et al.* studied the assembly at 1 M ionic strength whereas we carried out SAXS experiments at 0.5 M ionic strength. The kinetics in our experiments should be slower since the Debye screening length

was increased by a factor of $\sqrt{2}$. With our SAXS data, we did not see any traces of PODs, possibly because the reaction was too fast to dissociate between the early intermediates and/or because the signal predominantly came from the largest objects in solution. We indeed found instead an intermediate of the size of a half-capsid (35-45 dimers), which is probably an average of intermediates of similar size. Intermediates of such size were however not reported by Zlotnick *et al.* in their assembly studies.

Zlotnick *et al.* also found that a two-step reaction (nucleation-elongation mechanism) was relevant to model the self-assembly. In their model, the second step was the rate-limiting step. No matter how fast the first step took place, the overall reaction could not proceed any faster than the second step in the reaction. With the spectral information and the four kinetics at different initial concentrations that we possessed, we found out that this model was unlikely appropriate. Instead, a more straightforward kinetics model that also involved a two-step reaction, the second step not being an one-by-one addition of dimers, better fitted the data. Furthermore, the two steps of the reaction came into being strongly reversible as indicated by the values of the backward rate constants. This is contrary to the hypothesis of Zlotnick *et al.* established from the observations of the assembly species at equilibrium [34]. Indeed, the low abundance of such intermediates at the end of the kinetics would have signified that the capsid completion (second step) is an irreversible reaction. The small population of intermediates is rather explained in this mechanism by the high value of the backward rate constant of the first step which means that the activation energy of the reversible reaction (the one that produces dimers from intermediates) is low. The backward rate constant of the first reaction k_1^- is approximately 20 times higher than the backward rate constant of the second reaction k_2^-. This explains why there is a large population of dimers though this still depends on the total concentration. These remarks are illustrated with the calculated mass fractions (fig. 5-24) obtained from the fit with these values of rate constants.

As seen before, the fact that we could not identify species smaller than a half-capsid is due to the signal of the latter which is predominant. It is also because their lifetime was so short that they were not able to be discriminated. In contrast, the fact that species larger than the half-capsid object were not detected by TR-SAXS should signify that their typical lifetime is short, or at least shorter than the half-capsid intermediate. In other words, in the self-assembly process, the half-capsid intermediate appears to be the largest intermediate with the longest lifetime.

The observation of large objects in the assembly process of viral particles is not however completely new. Tuma *et al.* [99], who studied the assembly of the bacteriophage P22 procapsid, detected an intermediate which corresponded to about 30-40% of the capsid volume, which is the similar size of what we found with CCMV empty

capsids. The assembly kinetics was reconstructed from the Singular Value Decomposition (SVD) analysis and rotation of the eigenvectors. The radius of gyration of the detected intermediate was estimated using a model of an spherical cap.

5.6 Conclusions on Chapter 5

CCMV empty capsids assembly has already been studied by Zlotnick *et al.* in 2000 [24]. They concluded that the formation of capsids occurs with the initial formation of pentamers of dimers (PODs) that they observed with light scattering. This chapter re-examined the self-assembly of CCMV empty capsids by probing its kinetics by Time-Resolved Small-Angle X-ray Scattering (TR-SAXS). Using the TR-SAXS technique, we intended to provide insights into the self-assembly by retrieving structural information as the kinetics occurs.

The concept of this project was somehow quite similar to Chapter 4 when studying the disassembly of empty capsids by TR-SAXS. The objectives were to extract the form factors of the intermediates that could be detected as well as to devise a kinetic model that fitted the data. However, with the self-assembly study, we used a different approach that consisted of modeling the intermediates with an analytical object. The mathematical model consisted of a piece of capsid resulting from a stereographic projection of an ellipse onto the capsid. The form factor of this object was computed using a Fourier transform on a spherical harmonics decomposition. The method was numerically integrated into a more complex algorithm that the one used in Chapter 4. This is because we had to characterize a more complex system of data coupling of four sets of kinetics data at different concentrations into a single set of data. The resulting algorithm enabled us to extract the mass and the global shape of the only intermediate that could be detected in the four assembly reactions. The latter was found to be of the order of a half-capsid, which is far bigger than the size of the pentamer of dimers (PODs) that was detected by Zlotnick *et al.* with light scattering. Whereas Zlotnick *et al.* claimed that PODs were identified but did not report any half-capsid intermediates, we showed that PODs were not detected by TR-SAXS but half-capsid intermediates were observed instead and shown to be part of the assembly reaction.

We recently reported [89] the detection of an other half-capsid size object in the disassembly kinetics of CCMV VLPs, which supports our findings in the self-assembly process. Whether the two half-capsid intermediates had the exact same mass and structure was still uncertain. The average mass was partly established and was about of 30-40 dimers for the big intermediate detected in the disassembly process and of

about 35-50 dimers for the one detected in the self-assembly process.

We anticipated that the study of rapid kinetics involving unknown order of reactions must have taken into account the study of several kinetics at different initial concentrations. The absence of information concerning the structures of the intermediates have been compensated by a model of the reaction intermediates which constrained conveniently the kinetic model. Multiple data were found to constrain our kinetic model with additional constraints from the spectral aspect that was not provided by light scattering. We emphasized on the fact that a single kinetics model was used to fit four kinetics at four different concentrations which made our numerical implementation robust enough to support our findings.

Chapter 6

Assembly of CCMV around single-stranded RNA

In a seminal work, Fraenkel-Conrat [100] showed that active tobacco mosaic virus can be reconstructed from its inactive protein and nucleic acid components. The activity of the virus seems to simply lie in the suitable interaction between the protein and the nucleic acid. This finding opens the door to many more unsettled questions, namely about the efficiency of the assembly, the strong interactions actually involved in this process and, above all, the pathway of assembly of proteins around the genome.

In the last two chapters, we investigated the assembly/disassembly of CCMV empty capsids, i.e capsids devoid of any nucleic acids inside. Only the protein-protein interactions were thus investigated and the kinetics was only controlled with the pH of the solution. Based on the results obtained from the self-assembly mechanism, we intended to perform the same kinetics analysis on the self-assembly of CCMV coat proteins (CCMV CPs) around single-stranded RNA (ssRNA). Since ssRNA is a charged polymer, electrostatic interactions are now highly involved and the kinetics also has to be controlled by the ionic strength of the solution. The question remains to know how the CCMV CPs interact with the RNA with the two control parameters that are the pH and the ionic strength of the solution.

In this chapter, we first present a short literature survey relative to the assembly of viral proteins around a genome. Theoretical studies concern assembly around a genome and a polymer whereas the experimental works that are mentioned are only limited to on the genome encapsulation. We particularly focus on the experimental works that have been carried out by Gelbart *et al.* [101, 102, 103] and Zlotnick *et al.* [104] on CCMV CPs assembly around single-stranded RNA. Next, we present the experiments that we performed on CCMV CPs on the kinetics of encapsidation of their own RNA probed by TR-SAXS. The experimental setup and the methods utilized are exposed. Few analysis on TR-SAXS data are given together with cryo-TEM images for supporting the results.

6.1 Literature survey - Theoretical works on dynamics of assembly of viral proteins around a polyelectrolyte

6.1.1 Dynamics of assembly around a polymer: computational simulations

The dynamics of assembly around polymers was notably investigated with computational models. The model of a linear polymer particularly made the simulations more accessible.

Elrad and Hagan [105] developed a coarse-grained model to describe the assembly dynamics of T=3 empty capsids and cargo-containing capsids, in which the conformation of the subunit-subunit interactions was chosen to be varied. The simulations showed that only one morphology was identified in the assembly of empty capsids, but that different icosahedral morphologies were found in the assembly of nanoparticle cargoes of varying size.

In a different study, Elrad and Hagan [106] performed Brownian dynamics simulations of the encapsidation of a polymer. The subunit was modeled by a triangular form and the capsid geometry was a T=1. They particularly put forward two kinds of mechanisms in the assembly around the polymer by tuning both subunit-subunit and subunit-polymer interactions (see fig. 6-1). When subunit-subunit interactions are strong and subunit-polymer interactions are relatively weak, the assembly follows a nucleation-growth mechanism, where the initial formation of a small nucleus is followed by sequential addition of subunits. When subunit-subunit interactions are weaker and subunit-polymer interactions are strong, the capsid proteins bind to the polymer into a disordered complex, followed by a rearrangement and final capsid completion. Formation of a disordered micellelike structure followed by reorganization of subunits was also reported by McPherson [107].

Mahalik and Muthukamar [108] also explored the mechanisms of assembly into icosahedral structures using a linear polyelectrolyte with screened electrostatic interactions modeled by the Debye Huckel theory. They found similar results as those of Elrad and Hagan [106], that is to say two different mechanisms of assembly that depend upon the relative magnitude of protein-polymer and protein-protein interactions.

6.1.2 Assembly in the presence of nucleic acid: theoretical studies

The role of the nucleic acid in the assembly of viral particles is still intricate and puzzling. We present a non exhaustive list of theoretical studies that focused on the

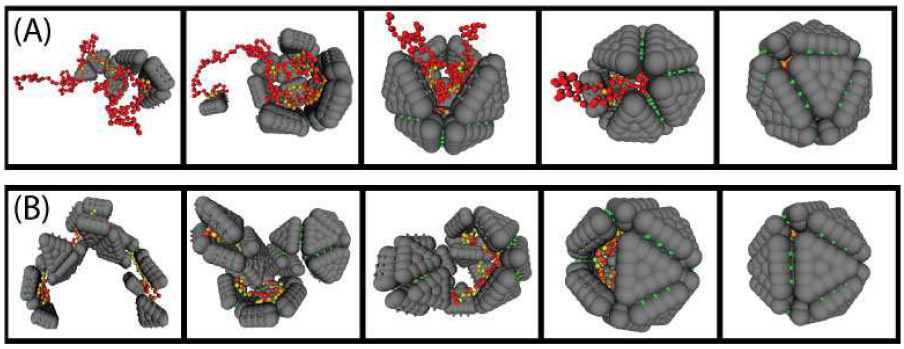


Figure 6-1: Two mechanisms for assembly around a polymer.**A**: Relatively strong subunit-subunit interactions and weak subunit-polymer interactions lead to a nucleation-growth mechanism. **B**: Relatively weak subunit-subunit interactions and strong subunit-polymer interactions lead to a disordered assembly mechanism. Image adapted from [106, 8].

specific role of the nucleic acid.

Hu *et al.* explored the role played by the Coulomb interaction between the positive N-terminal arms rooted at the inner surface of the capsid and the negative ssRNA molecule. They studied the Coulomb interaction in the capsid stability [109] and in the kinetics of *in vitro* virus self-assembly [110]. Zandi and Van der Schoot [111] established that the relationship between charge and volume of packaged nucleic acid was determinant for the particle size. Theoretical simulations of encapsidated dsDNA also underscore the importance of interaction with CPs in nucleic acid organization [112, 113, 114].

Von Hippel and co-workers proposed a model of nonspecific binding of proteins to a linear nucleic acid lattice [115, 116]. They described the association of one protein subunit with a nucleic acid substrate and the interaction between adjacent proteins was defined by the unitless association coefficient ω. Zlotnick and co-workers [117, 118] later expanded the theory of assembly accounting for the effect of nucleic acid conformation by introducing a second coefficient α. They introduced the coefficient α as a component of the assembly association constant. They also proposed a phase diagram of assembly based on the two parameters ω and α to put forward different regimes: low and high cooperativity assemblies and assemblies leading to kinetic traps and formation of aberrant structures.

6.2 Literature survey - Experimental works on the assembly of viral proteins with a genome

This section essentially comes out with some experimental works done on assembly of viral proteins with a genome. The first consists in investigating the optimal genome

length and CP:RNA mass ratio for total and efficient encapsulation. The second focuses on the assembly pathway from the co-assembly of RNA and CPs leading to a well-formed virus-like particle (VLP) from different conditions of pH and salinity.

6.2.1 Optimal genome length and RNA:CP mass ratio

"Magic ratio"

Cadena-Nava *et al.* [101] reported the assembly of BMV RNA and TMV RNA by CCMV CP (fig. 6-2). They first found that RNAs of very different lengths can be completely packaged as long as the protein:RNA weight ratio is sufficiently high, in this case 6:1 or greater. This finding was all the more surprising that it was independently of the length of the RNA. A comparable ratio was found in the *in vitro* packaging of CCMV RNA with CCMV CP, by Adolph and Butler [52] who reported a significant fraction of free RNA in their reassembly studies of CCMV involving a mass ratio of 5:1. They called "magic ratio" the value of 6:1, which they found

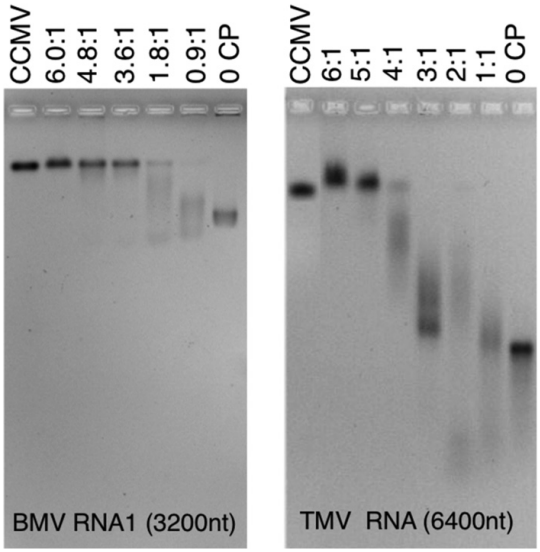


Figure 6-2: CP-RNA assembly titrations : gel retardation assays. *Left*: titration of BMV RNA1 with various amounts of CCMV CP ranging from 0 (right-most lane, RNA) to the "magic" ratio, 6:1 (second lane from left). CP/RNA ratios are wt/wt ratios of CP to RNA. *Right*: similar gel retardation assay carried out with TMV RNA. In both gels, the leftmost lane shows the position of wtCCMV. Image adapted from [101].

simply associated to a condition of charge equality between the RNA and the N-terminal arginine-rich motif of the CP (10 cationic residues for each 10 nt of RNA). This N-terminal is known to be the location of the protein where the RNA is binding to. Kobayashi and Ehara [119] in their *in vitro* encapsidation studies of the closely

related plant virus, cucumber mosaic virus (CMV), also found a 6:1 weight ratio for complete packaging of RNA. CMV CP N-termini also carry a charge of +10.

VLP products as a function of RNA length

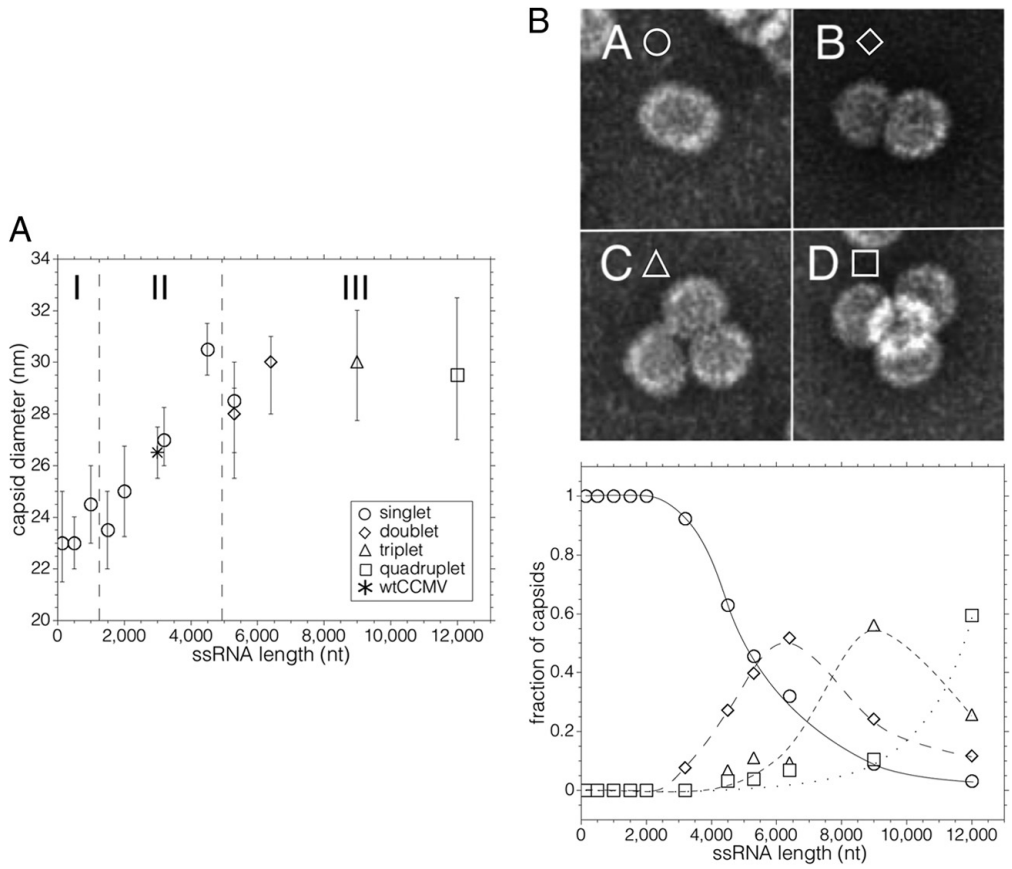


Figure 6-3: (A) Average capsid diameter as a function of RNA length. The plot is divided along the x-axis into three regions: (I) the region between 140 and 1,000 nt, where predominant singlet capsids are found; (II) the region between 1,500 and 4,000 nt also where singlet capsids are predominant; and (III) the region above 4,000 nt corresponding to the appearance of multiplet capsids. (B) *Top panel*: typical structures observed for singlets (A), doublets (B), triplets (C), and quadruplets (D). *Bottom panel*: Fraction of capsids present in the transmission electron microscopy as singlets (circles), doublets (diamonds), triplets (triangles), and quadruplets (squares). Image adapted from [101].

The 6:1 magic ratio for complete encapsulation of RNA by CCMV CP does not imply that all the products are capsids of an unique size, similar to the wild-type virus, i.e have a T=3 geometry. They actually found that, at the 6:1 mass ratio, the resulting products differ as a function of the RNA length. Figure 6-3 highlights the

formation of multiplets at high RNA length. Gelbart and co-workers [120, 63] also investigated the encapsulation of a polymer, poly(styrene sulfonate) (PSS) where they also found a capsid size distribution that strongly depended on the polymer molecular weight.

6.2.2 Insights into the assembly pathway of CCMV with RNA

Zlotnick *et al.* [104] carried out further studies on the assembly of RNA-containing CCMV VLPs. Their studies involved excess CCMV RNA and limited amount of CCMV CP under constant neutral pH conditions. With these conditions, they noted that the CPs initially bind RNA with low-cooperativity (after 20 mins), where the subunits are uniformly distributed over the many nucleic acid polymers (see figure 6-4). They reported a different mechanism that occurs later in the incubation (after 20-24 h), in which CPs slowly fold the RNA into a compact structure that can be bound with high cooperativity with the addition of CPs. The addition of CPs was found necessary and sufficient to promote the reorganization of the nucleoprotein complex to form the capsid. A stoichiometry of more than 90:1 (90 CP dimers for one molecule of RNA) which corresponds to a mass ratio of 4:1 (4 mg of CP per 1 mg of RNA), is required.

Gelbart *et al.* [102, 103] expanded their work by investigating the co-assembly of CCMV CP/BMV RNA with different pH. As seen in the last section, they found that a mass ratio of 6:1 is required for complete encapsulation of RNA. They demonstrated that the *in vitro* co-assembly of CP/RNA into a complete well-formed icosahedral capsid follow a two-step, pH dependent, mechanism. In the first step, CPs reversibly bind to RNA at neutral pH and form a disordered nucleoprotein complexes. Then, upon acidification, thess complexes undergo a structural reorganization to form the icosahedral capsid. They also showed that the spherical particles maintain their structure after dialysis back to neutral pH (7.2,0.1) buffer (see figure 6-5).

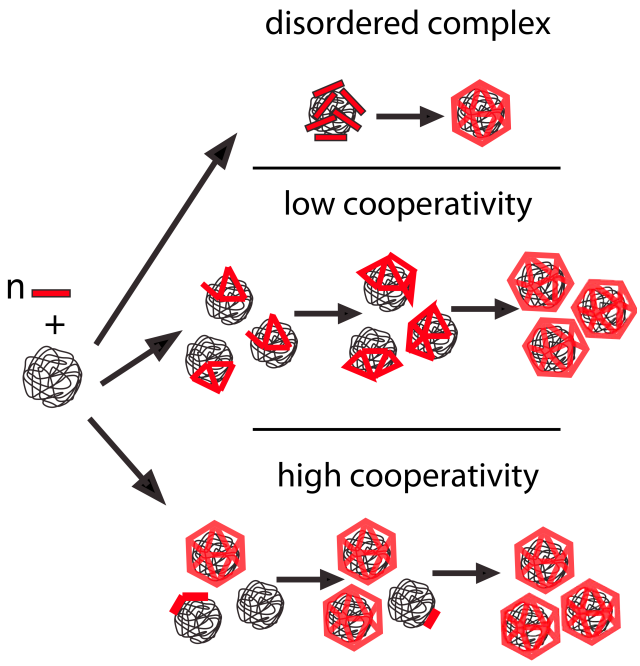


Figure 6-4: Three general pathways of assembly. *Top*: Subunits adsorb to the nucleic acids to form a disordered complex and then rearrange to form a capsid; this path has been associated with weak CP-CP interaction [121, 118, 103, 8]. *Middle*: A low-cooperativity assembly path has a more ordered addition of subunits, where the subunits are uniformly distributed over the many nucleic acid polymers. *Bottom*: A high-cooperativity path, also based on ordered addition of subunits, results from preferential addition of subunits to nucleic acid where growth is already nucleated. Low and high cooperativity assembly paths have been associated with strong CP-CP interaction and weak CP-nucleic acid interaction [118, 8]. Image adapted from [118].

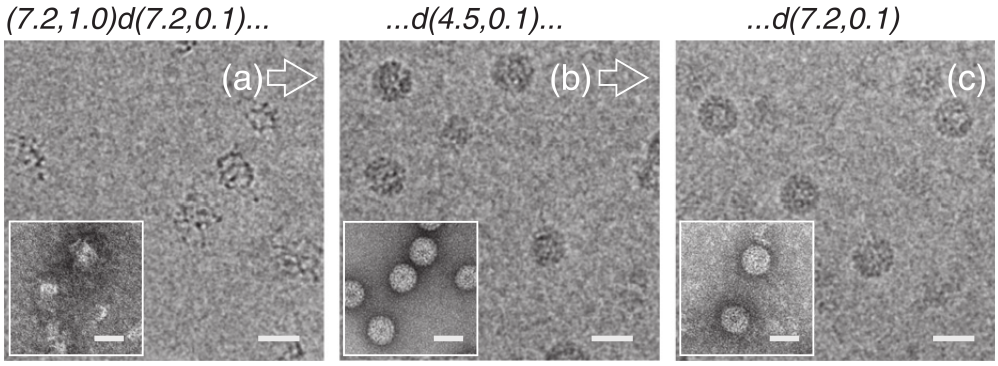


Figure 6-5: Cryo-EM micrographs (negative-stain images are shown in insets) of assembly reactions at different buffer conditions. (a) Images of disordered CP-RNA complexes at pH 7.2 and I=0.1 M denoting by (7.2,0.1). (b) Spherical VLPs formed by dialyzing the complexes in (a) against low pH (4.5,0.1) buffer. (c) The spherical particles formed in (b) maintain their structure after dialysis back to neutral pH (7.2,0.1) buffer. Scale bars represent 28 nm. Image adapted from [103].

6.3 Design of TR-SAXS experiments on CCMV CP/CCMV RNA assembly

6.3.1 Introduction

The results from Gelbart *et al.* are interesting because they demonstrated they could control the assembly into well-formed capsids only by tuning the pH and the ionic strength. Indeed, the low ionic strength (I=0.1 M) enabled the CP to bind the RNA [103]. Then, the acidification of the solution (pH 4.8) allowed the restructuration of the nucleoprotein complex into the capsid.

We decided to bring further insights into the assembly pathway of CCMV with RNA by investigating its kinetics by TR-SAXS. So far, the only work about following by TR-SAXS the assembly of viral proteins with nucleic acids that we are aware of, is the work done by Raviv *et al.* [38]. They focused on a T=1 SV40 VLPs made up of 12 identical subunits (pentamers of dimers) that encapsulated short single-stranded yeast RNAs (up to 800 nt). In our case, we proposed TR-SAXS experiments to follow the kinetics of CCMV (T=3 geometry) encapsulating its *own* RNA (up to 3,000 nt).

6.3.2 Experimental setup

6.3.2.1 X-ray scattering

The experiments were done in September 2014 at the ESRF (ID02 beamline). SAXS patterns were recorded using a high speed and ultra-low noise area CCD detector (Rayonix). The sample-to-detector distance was set to 3.0 m which provided scattering wavenumbers q ($q = 4\pi/\lambda \sin(\theta/2)$ where θ is the scattering angle and $\lambda = 0.996$ Å the X-ray wavelength) ranging from 2.4×10^{-3} to 0.26 $Å^{-1}$. The experimental setup included a stopped-flow apparatus (Biologic SFM-400) used to mix the CCMV CP with RNA at different buffer conditions (pH 7.5 or pH 4.8). The mixed solution was then injected through a quartz capillary of diameter 1.34 mm. The flow was stabilized with a hard stop valve and the temperature was maintained at 10°C via an external cryo-bath.

6.3.2.2 Mixing dead time

The time appended to each spectrum and the mixing dead time were calculated knowing the dead volumes and the total flow rate.

The dead volume between the syringe S1 and the last mixer was 50 μL. The dead volume between the last mixer and the capillary was 15 μL. The dead volume to reach the half capillary where the x-ray beam was positioned was 7μL. Hence the total dead volume of 72 μL. Because the flow rate was set to 8 mL/s, the time for the solution to reach the position of the x-ray beam from S1 was estimated to be 9 ms. The trigger,

which corresponds to the time between the beginning of the injection (solution in S1) and the first acquisition was set to 11 ms. The mixing dead time is considered to be the time when the solution is mixed and the information not captured during this time. In this case it is about 5 ms, since it corresponds to the time the solution reached the half capillary from the last mixer, plus the 11-9=2 ms dead time.

6.3.2.3 Preparation of samples and buffers

Buffers

We simply list the buffers used for the experiments:
- Buffer A : 50 mM Tris-HCl, 0.1 M NaCl, 1mM EDTA, pH 7.5
- Buffer B : 50 mM sodium acetate , 0.1 M NaCl, 1mM EDTA, pH 4.8
- Buffer C : 50 mM Tris-HCl, 0.5 M NaCl, 1mM EDTA, pH 7.5

CCMV coat proteins

CCMV coat proteins were purified using the same usual protocol and dialyzed overnight against the buffer A. The concentration was checked by absorbance and was c = 4.17 mg/mL. The absorbance ratio was $A_{280}/A_{260} = 1.47$.

CCMV ssRNA

The ssRNA was isolated from purified CCMV virions using the TRIzol® LS Reagent from Ambion. The RNA isolation protocol was adapted from the protocol furnished from the TRIzol® kit. The protocol was following appropriate precautions to avoid RNase contamination by sterilizing the equipments.

Dialysis at pH 7.5: CCMV virions were dialysed overnight against buffer C to enhance the swelling to favor the RNA extraction.

Homogenizing samples: A ratio of 3:1 was always maintained between the volume TRIzol® LS Reagent (TLSR) and the volume of the sample. To work with small quantities, 0.75 mL of TLSR was added per 0.25 mL of sample volume. The sample was then homogenized by pipetting the suspension up and down for 5 minutes.

Phase separation: 0.2 mL of chloroform per 0.75 mL of TLSR used for homogenization was added and the solution was shaked vigorously for 15 s followed by 15 mins of incubation at room temperature. The sample was then centrifuged at 12,000 × for 15 mins at 4 °C. The colorless upper aqueous phase was pipetted out and placed in a new tube.

RNA precipitation: 0.5 mL of 100% isopropanol was added to the aqueous phase, per 0.75 mL TLSR used for homogenization. The sample was incubated at room temperature for 30 mins then centrifuged at 15,000 × for 15 mins at 4 °C.

RNA wash and resuspension: The supernatant was removed, leaving only the RNA pellet. The latter was washed with 1 mL of 75% ethanol per 0.75 mL of TLSR used for homogenization and the sample was mixed vigorously for 5 mins. The sample was

then centrifuged at 9,000 × g for 10 mins at 4°C and the supernatant was discarded. The RNA pellet was air dried for 20-30 mins and resuspended in sterilized water in sterilized tubes by pipetting the suspension up and down for 5 minutes. The sample was then stored at -80°C after checking its absorbance and concentration.

For the TR-SAXS experiments, the ssRNA sample concentration was 0.68 mg/mL and the absorbance ratio were $A_{260}/A_{280} = 2.14$ and $A_{260}/A_{230} = 2.40$.

6.3.3 TR-SAXS experiments: General procedure

We carried out experiments with SAXS based on the methodology of Gelbart *et al.*[103] by investigating the kinetics of the one-step assemblies and two step-assemblies as described below. The general procedure simply consists of mixing CP with RNA at the 6:1 mass ratio at the desired values of pH and salinity. The CCMV CP were initially dialyzed against buffer A (pH 7.5, 0.1 M NaCl). The RNA was separately dialyzed against the same buffer A.

One-step assembly

The one-step assembly consisted in mixing three solutions : CCMV CP solution, RNA solution and buffer B (pH 4.8, I=0.1 M), at the same 6:1 mass ratio and varying the total concentration (see Fig. 6-6).

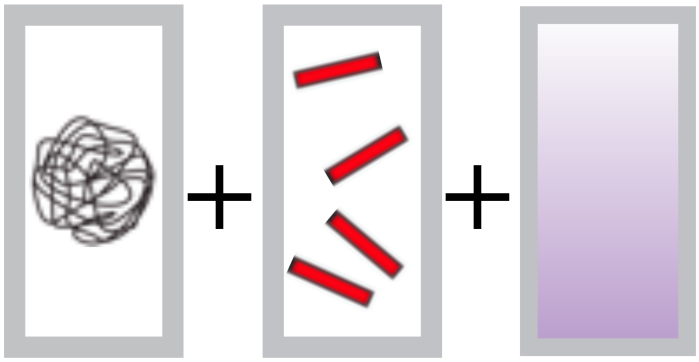


Figure 6-6: One-step assembly: mixing simultaneously three solutions. RNA is represented in dark line, CP in red blocks and acidic buffer in purple.

Two-step assembly

The two-step assembly consisted first in the pre-assembly of CCMV CP and RNA (CP-RNA binding) at 6:1 mass ratio by dialyzing overnight in buffer A (pH 7.5, I = 0.1 M). A nucleoprotein complex is formed. The second step was the acidification of the sample to a final state (pH 5.2, I = 0.1 M) by mixing with buffer B. This is illustrated in Fig. 6-7.

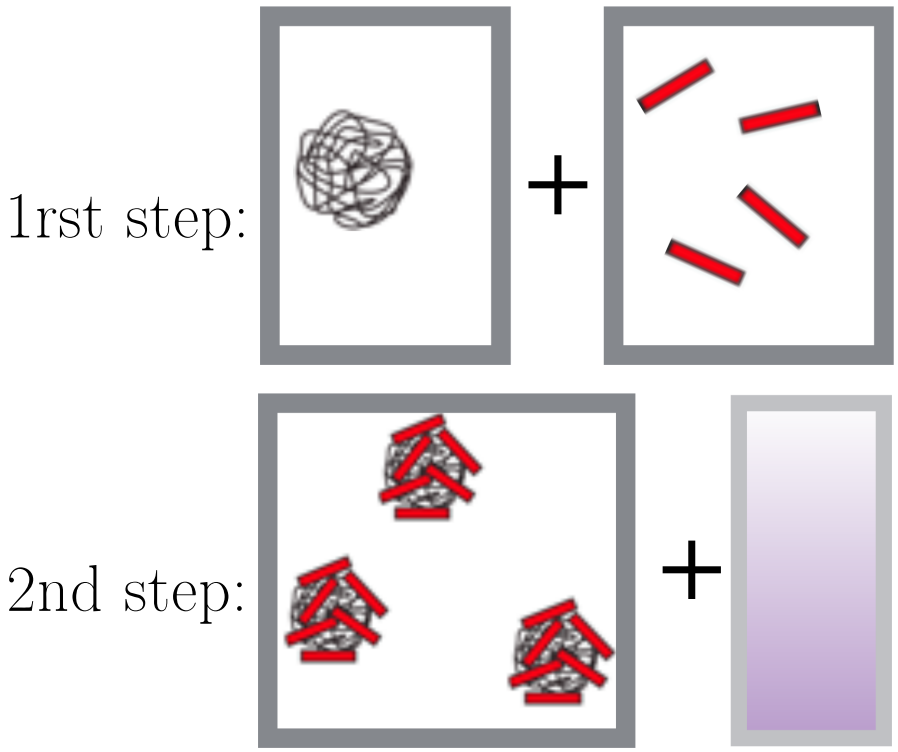


Figure 6-7: Two-step assembly: (*Top*): CP-RNA binding at neutral pH leads to nucleoprotein complexes. (*Bottom*): The second step is an acidification step. RNA is represented in dark line, CPs in red blocks and acidic buffer in purple.

6.3.4 Cryo-TEM images

In parallel to the TR-SAXS experiments, we used the cryo-microscopy technique to visualize if similar assembly products at equilibrium as mentioned by Gelbart *et al.* are obtained with the assembly of CCMV CP with its own RNA (CCMV ssRNA) (see figure 6-8).

Protocol

CCMV coat proteins were purified with the usual purification protocol. CCMV ssRNA was purified with the above protocol. CCMV CP concentration was 6.2 mg/mL and CCMV ssRNA concentration was 1.1 mg/mL so that a 6:1 mass ratio was respected. CCMV CP and CCMV ssRNA were mixed and dialyzed overnight against buffer A. Half of the resulting solution was laid aside for cryo-microscopy. The other half was mixed with equal volume of buffer B. The two samples were then diluted so that the final protein concentration were 0.3 mg/mL.

Images

We carried out TEM experiments on the two-step assembly, that consisted in mixing

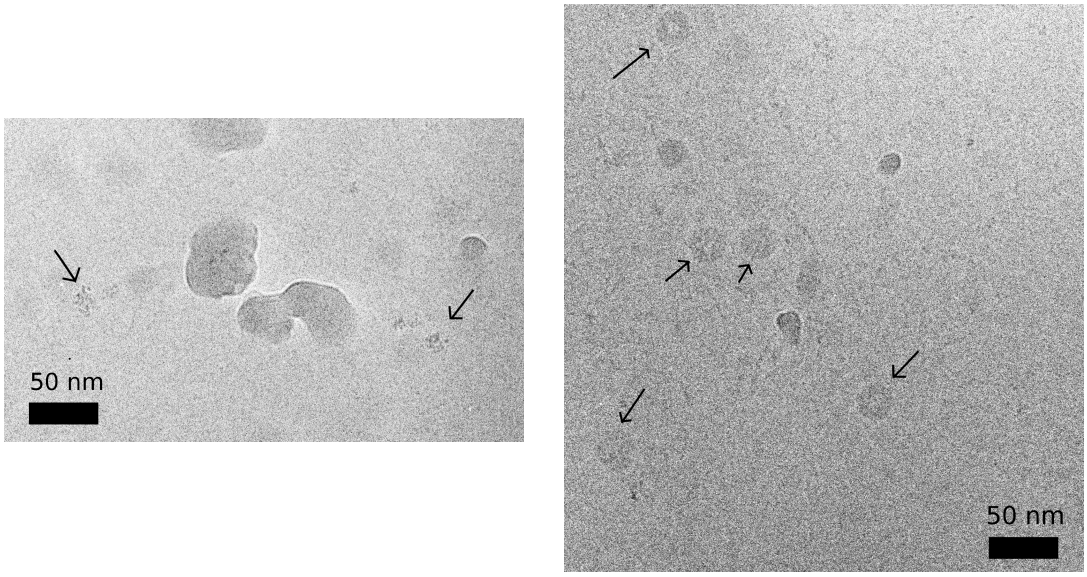
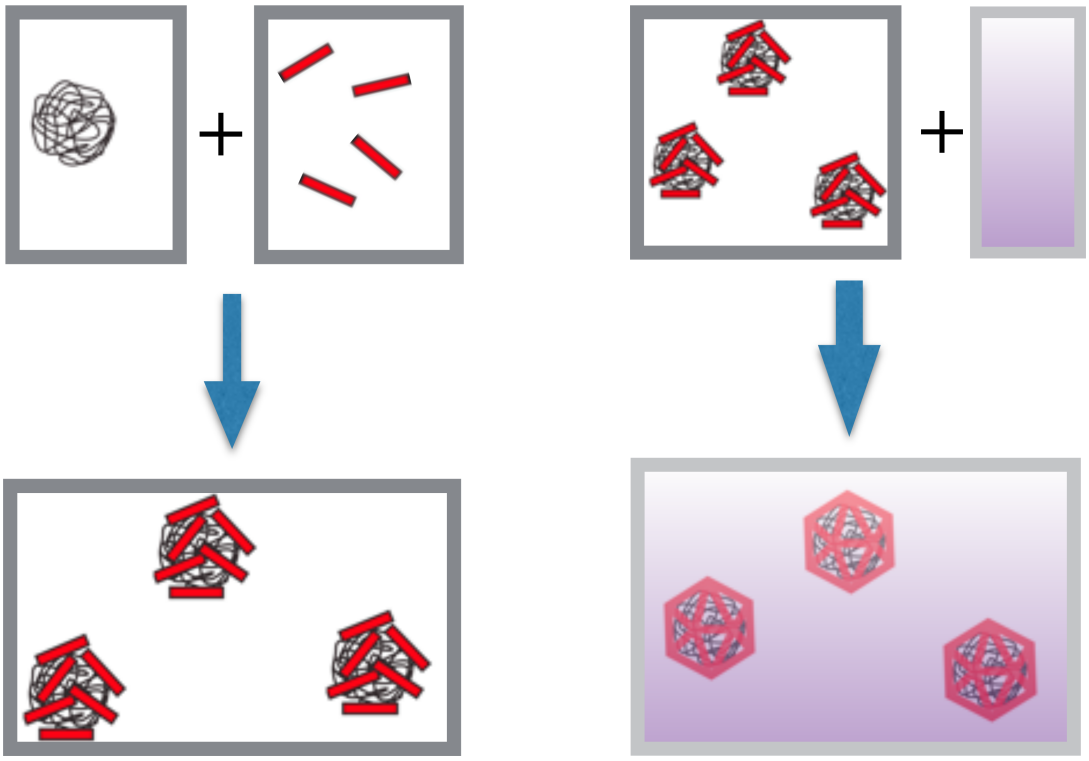


Figure 6-8: Cryo-TEM images of assembly of CCMV CP with CCMV ssRNA at different buffer conditions. *Left*: CP-RNA complexes obtained from dialyzing in (pH 7.5, I=0.1 M) buffer. *Right*: Most particles formed by dialyzing the complexes against low pH (pH 5.25, I=0.1 M) were spherical VLPs. There was still a few population of disordered aggregates of CP-RNA complexes. In both images, the gray blots were ice crystals and the arrows indicate the particles from the reaction products.

CCMV CP with CCMV ssRNA at pH 7.5 and ionic strength I=0.1 M (CP-RNA binding), and in acidifying the resulting solution. At pH 7.5 and ionic strength I=0.1 M,

the reaction products formed were found to be aggregates, disordered nucleoprotein complexes (Fig. 6-8 *Left*). At pH 5.25 and ionic strength I=0.1 M, most particles were spherical (Fig. 6-8 *Right*). We noted however the presence of few disordered nucleoprotein complexes as well. The results are illustrated in Fig. 6-9.



(a) 1rst step (b) 2nd step

Figure 6-9: Two-step assembly. (a) 1rst step: CP-RNA binding at neutral pH leads to the formation of nucleoproteic complexes. (b) 2nd step: The acidification step makes most complexes reorganized in spherical objects. RNA is represented in dark line, CPs in red blocks and acidic buffer in purple.

With this low value of ionic strength, the electrostatic binding of CP-RNA is turned on and CPs primarily interact with RNA through a disordered motif. Residues 1-26 of the CCMV CP contribute +10 charges that interact strongly with the negatively charged phosphate backbone of the RNA [103]. The decrease in pH should enhance the hydrophobic and electrostatic interactions between proteins, leading the the final completion and reorganization into spherical particles. The attraction between CPs have been shown to be pH dependent, increasing with decreasing pH [34]. This is due to electrostatic attractions resulting from the protonation of acidic residues at acid pH [122].

6.4 Presentation of the TR-SAXS data

In this section, we present some TR-SAXS data collected during the synchrotron experiments. The first set of data deals with the mixing of CP and RNA simultaneously with a pH jump (one-step assembly). The second set of data concerns the kinetics CCMV CP/RNA binding at pH 7.5 and ionic strength I=0.1 M. The kinetics of acidification of pre-formed nucleoproteic complexes is not mentioned as few doubts arose from the SAXS experiments concerning the mixing with a buffer to acidify the solution or concerning the pre-assembly dialysis performed few days before the SAXS experiments.

6.4.1 Mixing CCMV CP and CCMV ssRNA with a pH jump - pH 5.25, I=0.1 M (One-step assembly)

We performed experiments on the mixing of CP and RNA simultaneously with a pH jump (one-step assembly). Three kinetics at three different concentrations were collected (0.5, 0.7 and 1 mg/mL). All of these three kinetics systematically showed aggregation (only one concentration is plotted in figure 6-10). Aggregation might happen because the system may be kinetically trapped due to the enhanced CP-CP attractions because of the decrease of pH. The shape of the final spectra however suggest that the same kind of particles are formed, that is, irregular capsids. Gelbart *et al.* [103] also visualized aggregates of irregular capsids with electron microscopy when investigating the CCMV CP-BMV RNA mixing with a pH jump (one-step assembly).

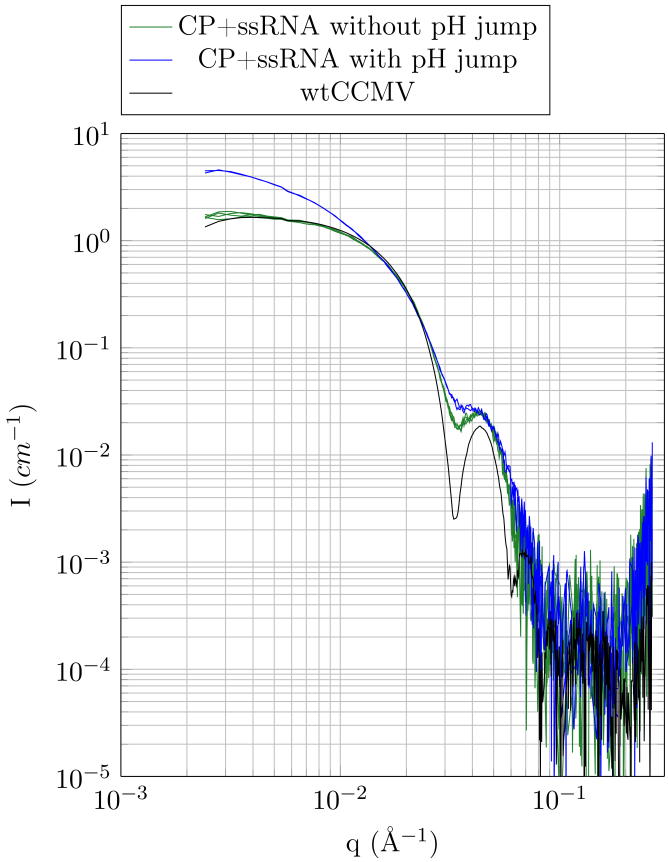


Figure 6-10: One-step assembly: CCMV CP and CCMV ssRNA with pH jump -SAXS patterns obtained from different experiments. The green spectra represent the four final spectra of the kinetics of CP-RNA mixing without pH jump at 1 mg/mL. The radii of gyration was computed by AUTORG: Rg=110±4 Å. The experimental SAXS pattern of the purified wtCCMV (black) has a Rg=100±6 Å. The final SAXS patterns of the kinetics of CP-RNA mixing with pH jump are at 1 mg/mL (blue curves) and showed strong aggregation. The intensities were in absolute units except for the black curve which has been superimposed to the green curves to enable comparison.

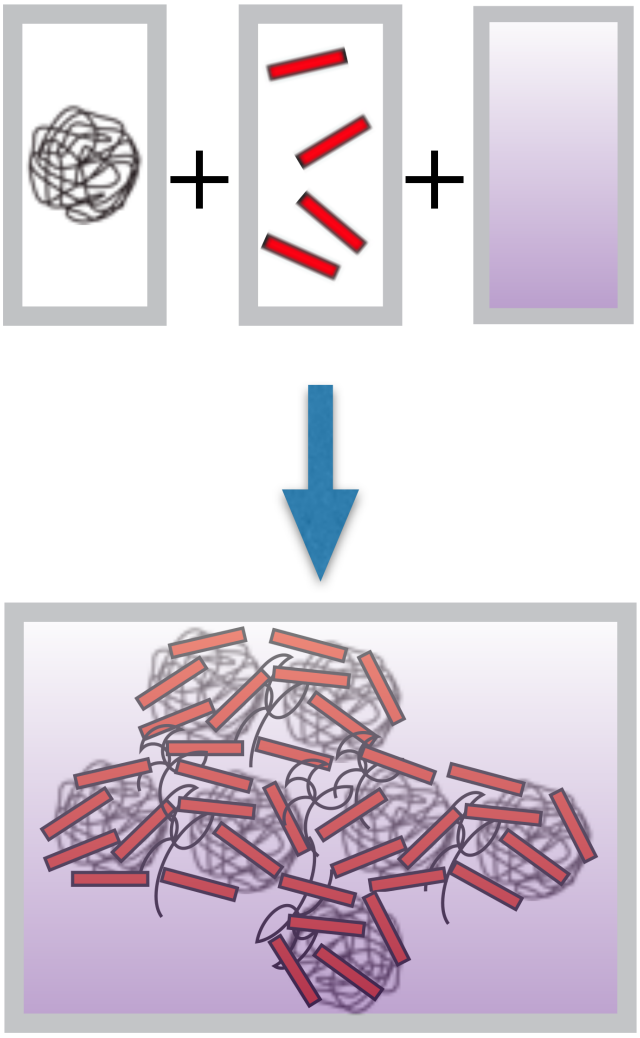
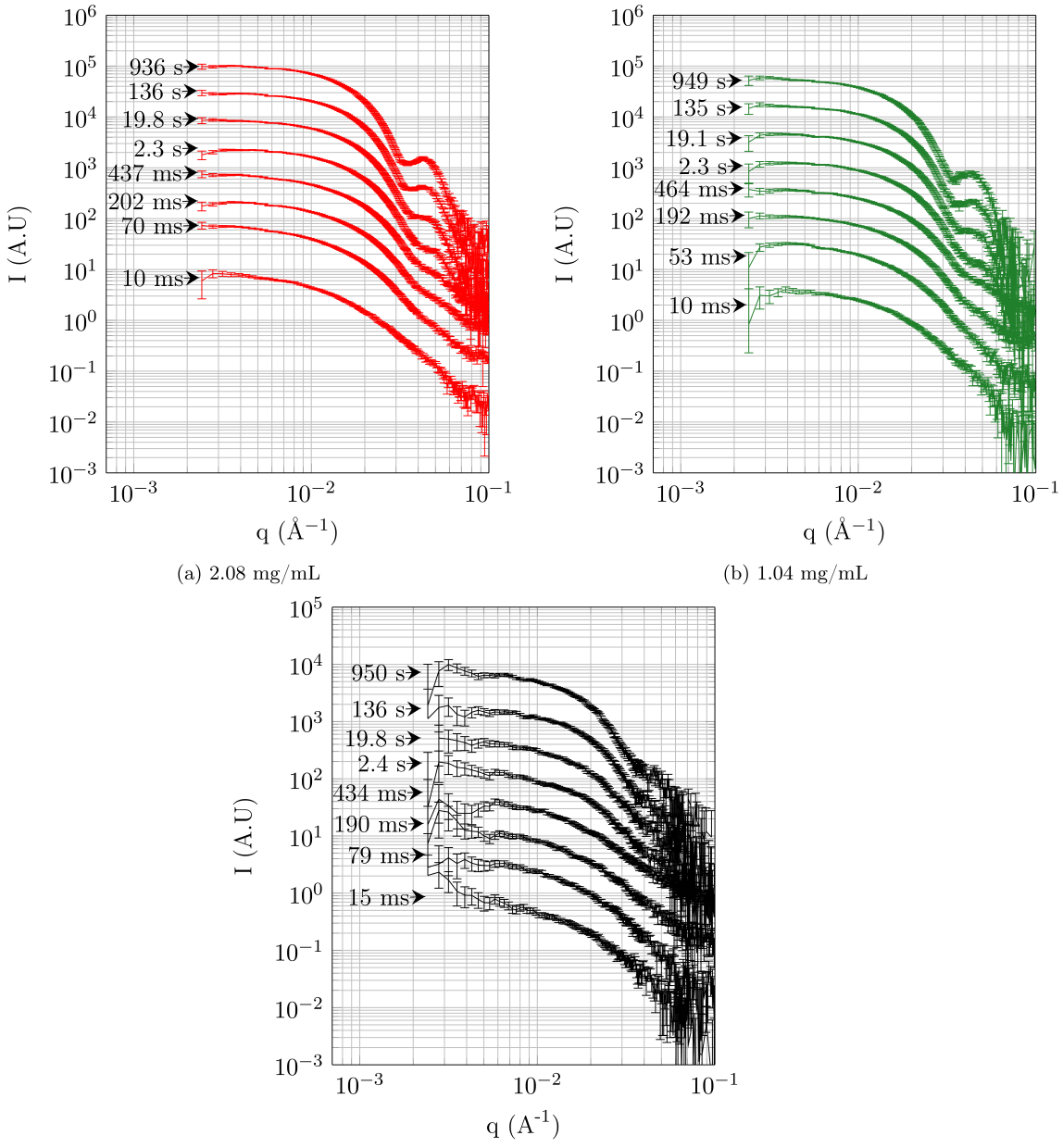


Figure 6-11: One-step assembly - Results from TR-SAXS experiments. Mixing simultaneously three solutions gives aggregation of complexes of irregular shape. RNA is represented in dark line, CP in red blocks and acidic buffer in purple.

6.4.2 CP/RNA binding. Mixing CCMV CP and CCMV ssRNA at neutral pH (pH 7.5, I=0.1 M)



(a) 2.08 mg/mL

(b) 1.04 mg/mL

(c) 0.52 mg/mL

Figure 6-12: TR-SAXS patterns obtained from the assembly of CCMV CP with CCMV ssRNA at pH 7.5. The patterns are shifted for clarity. The concentrations indicate the protein concentrations. The CP:RNA mass ratio was 6:1. The difference in the time indexation between the three kinetics is due to a change in the time acquisition.

We later performed TR-SAXS experiments on the kinetics of CP/RNA binding

at neutral pH to observe the formation of nucleoproteic complexes (see 6-9 (*Left*)). Figure 6-12 shows several TR-SAXS patterns collected at different times of the kinetics for three different concentrations. The difference in the time indexation between the three kinetics is due to a change in the time acquisition. The superimposition of the final spectra for the three kinetics show that the reactions are finished in each case. The final patterns are shown together with the wtCCMV pattern in figure 6-10. The radii of gyration were computed and were found to be similar within the errorbars, which suggests that the final products are comparable in size and shape to the wild-type viruses.. Whether the real difference in the patterns from the region 3×10^{-2} - 1×10^{-1} Å^{-1} is due to a distinct particle or due to the presence of free RNA and CP in the solution is not completely established yet, although the TEM images revealed that these are most probably distinct particles. Perhaps the more probable explanation is that, in the final state, there is the presence of both population of well-formed capsids and disordered nucleoprotein complexes at lower concentration, which explains why the radii of gyration are close to the radius of gyration of the well-formed capsid and why the typical oscillations of the form factor are not well pronounced.

At this stage, it seems reasonable to conduct more SAXS experiments at higher angles q, at equilibrium, and possibly with varied CP:RNA mass ratio and total concentration. It is indeed difficult to model the form factor of such nucleoprotein complexes that look disordered from the TEM images. It may be possible to filter the solution to retain only the nucleoprotein complexes and discard the free RNA and CP. The SAXS spectrum of such complex may be in this case be determined experimentally.

6.5 Forward scattering intensities and assembly rates

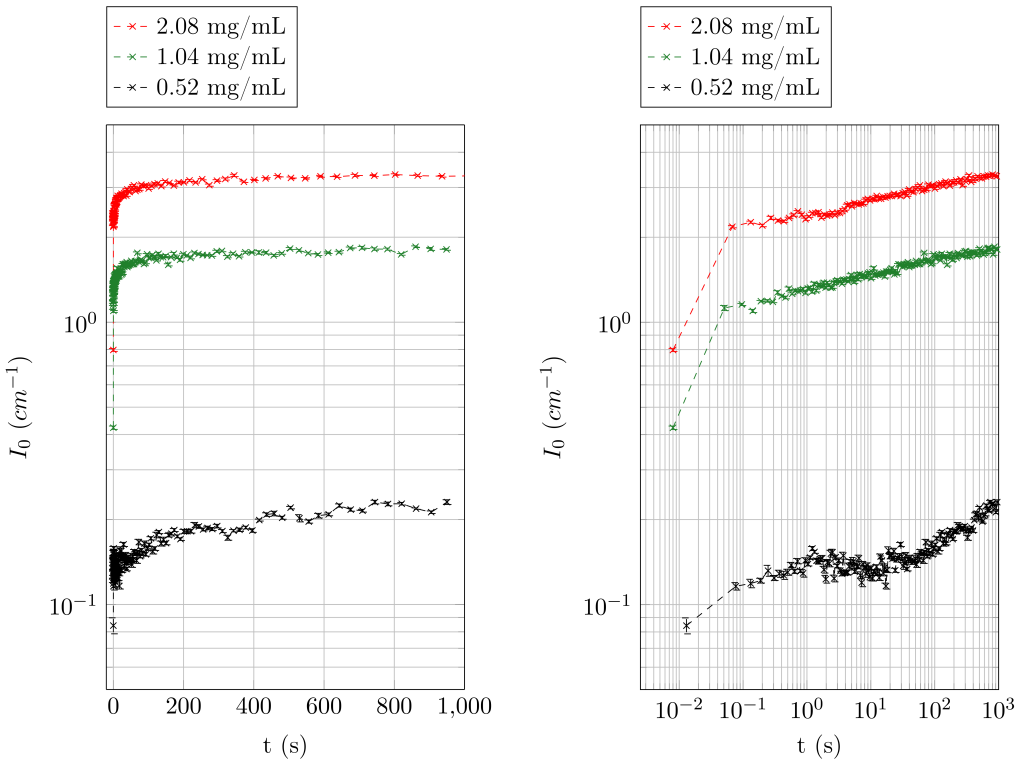
Analysis on data obtained by mixing proteins and RNA with a pH jump to a final pH of 5.25 is discarded because of the apparent aggregation observed in the kinetics. Figure 6-13 displays the forward scattering intensities I_0 as a function of time for the kinetics of CCMV CP-RNA mixing without pH jump.

One can estimate how fast is the kinetics by looking at the assembly rate at the very early times of the reaction. The (normalized) assembly rate ν (in s^{-1}) is indeed calculated by

$$\nu = \frac{1}{I_0(t=0)} \frac{dI_0}{dt}\bigg|_{t=0} \tag{6.1}$$

Table 6.1 shows the values of the assembly rates estimated during the first 0.5 s of the acquisition.

To enable comparison with the assembly of CCMV empty capsids, we also calculated their assembly rates (during the first 3.5 s of the acquisition) which are rep-



(a) lin-log scale

(b) log-log scale

Figure 6-13: Forward scattering intensities I_0 as a function of time computed by AUTORG for the kinetics of CCMV CP-RNA binding at neutral pH represented in linear-logarithmic scale (a) and logarithmic-logarithmic scale (b).

$Concentration(mg/mL)$	ν (s^{-1})
0.52	1.09
1.04	3.78
2.08	3.87

Table 6.1: Assembly rates estimated from the first 0.5 s of the acquisition from the kinetics of mixing of CCMV CP and CCMV ssRNA. The pH and ionic strength were kept constant during mixing (pH 7.5 and I=0.1 M).

$Concentration(mg/mL)$	ν (s^{-1})
0.48	0.07
0.67	0.27
0.96	0.44
1.49	0.75

Table 6.2: Assembly rates estimated from the first 3.5 s of the acquisition from the kinetics of self-assembly of CCMV empty capsids. The final pH was 5.25 and the ionic strength 0.5 M.

resented in table 6.2. Clearly, the assembly of CCMV CP with RNA is much faster than the self-assembly of CCMV empty capsids, roughly by a factor of 10 if we look at the concentrations near 0.5 and 1 mg/mL. This was expected since the CP-RNA interactions are strong electrostatic interactions that should speed up the assembly reactions. The empty capsid assembly, on the other hand, is mainly driven by hydrophobic interactions, which make the assembly slower, especially if the proteins have to find the correct orientation to form spherical objects.

6.6 Radius of gyration

6.6.1 RNA spectrum

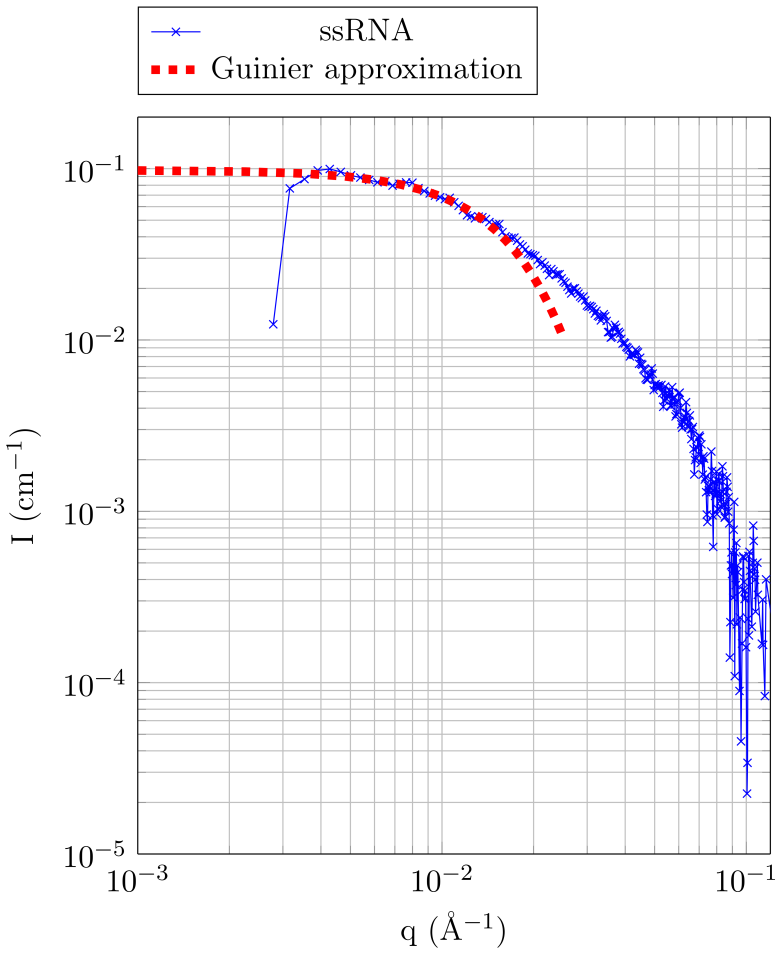


Figure 6-14: ssRNA experimental scattering pattern. The concentration was c_{RNA}=0.34 mg/ml in a buffer (pH 7.5, I=0.1 M). A Guinier fit is represented and the calculated radius of gyration was R_g=103 Å.

The scattering pattern of the ssRNA was recorded in Fig. 6-14. The Guinier

approximation resulted in the computation of the radius of gyration R_g=103 Å. The R_g of a Gaussian chain is given by R_g=$\sqrt{a^2N/6}$, where N is the number of monomers and a the Kuhn length. a is linked with the persistence length l_P by a=$2l_P$. Assuming the model of a Gaussian chain, the persistence length of the ssRNA being l_P=15 Åin 0.1 M NaCl [123] and N=3,000 nt, we then have $R_g^{(Gaussian)}$= 805 Å.

The ssRNA is compact but the experimental value R_g was found to be similar of the one of the CCMV empty capsid at acid pH (pH 4.8), which is ≈ 110 Å. The CPs are remarkably able to encapsulate a charged polymer of the same size as the final virus. We remind that Gelbart *et al.* [101] also reported the encapsulation of ssRNA of different length, ranging from 140 nt to 12,000 nt. CPs were found to encapsulate ssRNA into a capsid (singlet) for ssRNA length up to 5,000 nt. For higher length, the CPs encapsulated the ssRNA into structures of multiplets (see Fig. 6-3).

6.6.2 Average radius of gyration as a function of time

Fig. 6-15 shows the R_g of each scattering pattern of each kinetics at pH 7.5 and I=0.1 M presented in the last section using AUTORG [87]. It was interesting to observe that, as the kinetics occurs, the average R_g of the system was decreasing. It suggested that the CPs first bound to the RNA (hence a R_g ≈ 130 Å, which was higher than the R_g of the RNA only). The second process was that RNA was being compacted leading to the formation of a nucleoprotein complex. Furthermore, as the forward scattering intensity I_0 was still increasing as R_g was decreasing (up to ≈ 500 s), it suggested that the RNA was being compacted while additional CPs were binding to the RNA at the same time.

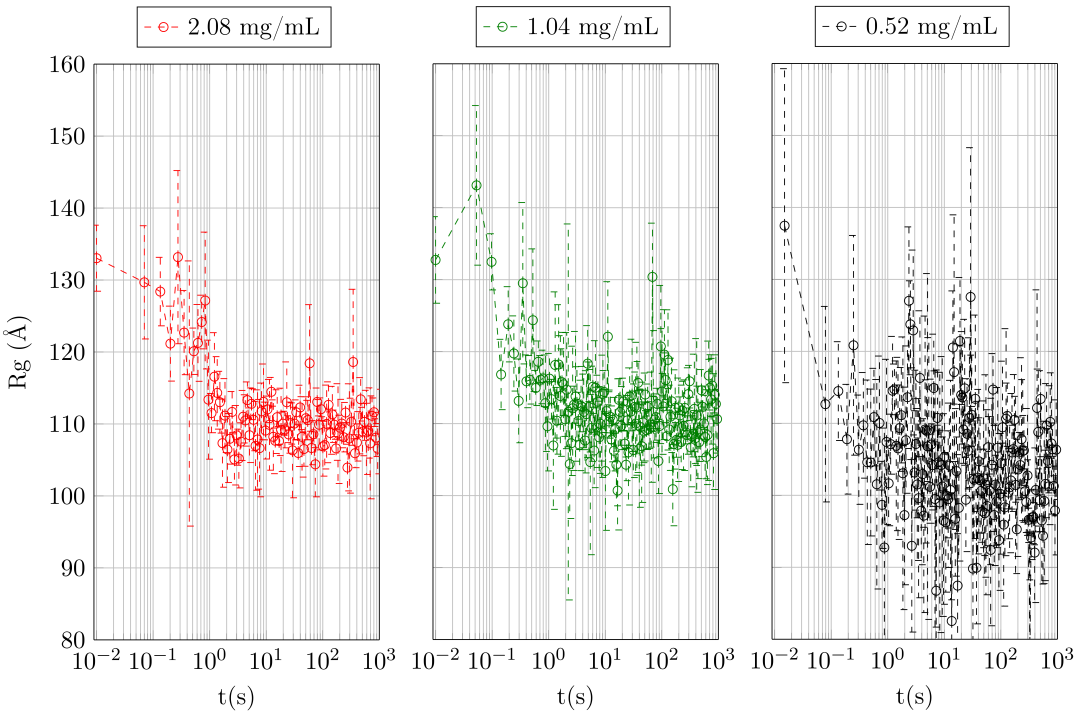


Figure 6-15: Evolution of the average radius of gyration. The radius of gyration of each TR-SAXS patterns is computed by AUTORG [87] for each of the three kinetics at pH 7.5 and I=0.1 M. The concentrations indicate the protein concentrations. The CP:RNA mass ratio was 6:1.

6.7 Conclusion on Chapter 6

In this chapter, we presented our project about the assembly of CCMV coat proteins around its own RNA followed by TR-SAXS. We referred to the work of Gelbart *et al.* who focused on the assembly products of CCMV coat proteins with a BMV RNA visualized by electron microscopy at the equilibrium. By investigating the kinetics with synchrotron source, we intended to bring further insights into the kinetics of assembly of CCMV proteins around nucleic acids. The structures of the reaction intermediates may not be resolved since the kinetics has been found to be much faster than the kinetics of empty capsids. The data however are expected to discriminate between the kinetics model that will be tested, as it was the case with the kinetics of assembly of empty capsids.

It was seen that the mixing of CP and RNA simultaneously with a pH jump systematically lead to aggregation of the particles. The system may be kinetically trapped due to the enhanced CP-CP attractions because of the decrease of pH. In the case where CP binds RNA in the same conditions of pH and salinity, this aggregation was not observed. This makes the modeling work on these data more achievable.

From the preliminary results we have, the mechanism of assembly of RNA through a disordered pathway predicted from computational simulations [106] and experimentally observed at equilibrium [103], seems to be accessible by TR-SAXS. There is still some work to do, namely to test different kinetics model on how CPs bind RNA and how fast they bind it. Additional SAXS experiments and microscopy should be needed in order to better characterize the nucleoprotein complexes formed during these reaction. This might be done by filtering the solution to retain only the nucleoprotein complexes. The pH might also be slightly varied (decreased to a final pH of 4.8) to better visualize the effect on the acidification of the solution.

Finally, the assembly of CP with RNA at pH 7.5 should be investigated as a function of the relative concentration of CP and RNA. I suggest to perform few kinetics at a mass ratio lower than 4:1 (stoichiometry of 90:1), according to the results from Zlotnick and co-workers [104] in order to double check and characterize the nucleoprotein complexes only with no well-formed capsids in solution. A second assembly at a mass ratio much greater than 6:1 (excess of CPs) should be carried out as well, to see if there are only well-formed capsids at the end of the kinetics. From what I have read from Zlotnick *et al.* and Gelbart *et al.*, there might be two ways to assemble CCMV CP with RNA into well-formed capsids. The first is to mix CPs and RNA at pH 7.5 and I=0.1 M at a mass ratio greater than 6:1 (excess of CPs). The second is to first mix CPs and RNA at pH 7.5 and I=0.1 M at a mass ratio lower 4:1 (excess of RNA) and operate a pH jump to a final pH=4.8.

Chapter 7

General conclusion and Outlook

This thesis presented the projects and the results obtained from the work I have done during my 3-year PhD in the Laboratoire de Physique des Solides. These projects examined different kind of kinetics of reaction with viral particles with the Time-Resolved Small-Angle X-ray Scattering (TR-SAXS) technique: the self-assembly kinetics of CCMV empty capsids (Chapter 5), their disassembly kinetics (Chapter 4) and the assembly kinetics of CCMV coat proteins with its own purified RNA (Chapter 6).

The manuscript first attempted to put those projects in a wide context of assembly into empty capsids or hybrid viruses. A large number of studies at equilibrium on these assemblies could be listed. However, few references could be quoted concerning the kinetics of assembly of viral protein subunits into well-formed, monodisperse spherical capsids. The self-assembly process can be studied theoretically, mostly with particle-based simulations or more recently with molecular dynamics simulations. Experimentally, it can only be scrutinized by extracting information during the assembly kinetics. The TR-SAXS technique was suited for the study of assembly of nanoscale objects since this technique allows one to get structural information as a function of time. When dealing with the TR-SAXS technique, one especially investigates the structure of the reactants, products and intermediates produced during the assembly reaction. This is possible due to the high brilliance of the synchrotron light and the recent development of stopped-flow techniques that enable to access the very early times of the kinetics.

We presented in this thesis, especially through the studies of the disassembly and self-assembly of CCMV empty capsids, a general method based on global fitting to reconstruct the kinetic pathway. We provided a kinetic scheme with the detected reaction intermediates that we estimated an average mass and structure.

In a broader context, the method used in the project on the disassembly kinetics of CCMV empty capsids (Chapter 4) was adapted from the paper published by Tresset *et al.*[27]. The project on the self-assembly of CCMV empty capsids (Chapter 5) used a different and new approach with the development of analytical objects to model the reaction intermediates. In Chapter 4, I used the global fitting method, that I adapted to the project, by improving the numerical implementation using suitable numerical algorithms. I also took into account larger probabilities on the expected size of the intermediates by testing several possibilities that respect the stoichiometry. The algorithm and its implementation had to take into consideration the fact that not one but two intermediates were detected during the reaction. The project ended with the results published in August 2015 (see [89]). In Chapter 5, the intermediate was modeled as a piece of capsid resulting from a stereographic projection of an ellipse. I numerically integrated the method into a complex algorithm to describe a complex system of data which is the coupling of the four sets of kinetic data at different concentrations. The resulting algorithm enabled us to extract the mass and the global shape of the only intermediate that could be detected. The latter was found to be of the order of a half-capsid, which is far bigger than the size of the pentamer of dimers (PODs) that was detected by Zlotnick *et al.*[24] with light scattering. Whether the half-capsid intermediate had the same mass and structure than the "big" intermediate occurring during the disassembly process was still uncertain but the comparison between the calculated mass made the results robust enough to support our findings. Indeed, the mass was partly established and was of about 30-40 dimers for the big intermediate detected in the disassembly process and of about 35-50 dimers for the one detected in the self-assembly process. As for the structure, one should bear in mind that the computed spectra correspond to an average, which should correspond to an average on the mass and on the spatial configuration. In the case of the self-assembly, the structure was modeled as a piece of a capsid that constrained the curvature of the intermediate, but the figures we showed illustrated well that there was still an uncertainty on the structural conformation. Nevertheless, two major results were found in this study. The first was that the capsids built up through the nucleation and the completion of a half-capsid. The second was that these two reactions (nucleation and completion) were strongly reversible.

In the self-assembly/disassembly of CCMV empty capsids, tuning the pH only enabled one to inspect only the role of the protein-protein interactions, which were mainly driven by hydrophobic and electrostatic interactions. The project we started on the assembly kinetics of CCMV CP with its own purified ssRNA was presented in Chapter 6. An additional complexity was added with the protein-genome interactions. The questions remained : how different was the kinetics of reaction? Were the same reaction intermediates found in this case? To start with this puzzling problem, the project strongly took into consideration the results obtained from Gelbart *et al.*

who thoroughly investigated the assembly of CCMV CP with BMV RNA at different conditions of pH and salinity. Although a different type of RNA was used, our prior investigations on the assembly CCMV CP/CCMV ssRNA by TEM suggested that the assembly products were similar from the ones observed by Gelbart *et al.*[103] at physiological pH, that is, disordered capsids. No SAXS experiments were previously investigated on these kinds of disordered nucleoprotein complexes resulting from one-step assembly. The SAXS data we presented intended to bring new insights on the process of assembly of this complexes.

The above experiments demonstrate the possibility of real time measurements with the TR-SAXS technique when the high brightness of the x-ray source is combined with a stopped-flow device. The TR-SAXS allows one to resolve the kinetic pathway of the self-assembly/disassembly mechanisms and to extract the overall structure of massive and long-lived reaction intermediates. It appears, from the conclusions made on these projects, that a better time acquisition seems to be crucial to investigate the very first steps of the reactions where smaller intermediates start to form. This may be done by improving the stopped-flow instrumentation, especially the mixing efficiency. But a better way to deal with this problem would be to find a way to trap these intermediates. Clarifications should be needed also concerning the mass of the reaction intermediates and nucleoprotein complexes. Native mass spectrometry experiments [29, 28] should be an ideal complementary technique to better understand the assembly and disassembly mechanisms, since the mass of the objects are provided with better accuracy.

The determination of the mass and the total number of intermediates are still open questions. But the methods we used in our study already bring more information concerning the speed of the kinetics and the overall mechanism with the presence of long-lived reaction intermediates. We anticipate that these methods can also be applied for a larger panel of reactions involving other supramolecular assemblies.

Appendix A

Quantification of RNA in CCMV capsids by SAXS

SAXS was used to determine the dimensions of the viral particles and to quantify the amount of encapsulated RNA. The study was done from the solution scattering of both wtCCMV and CCMV empty capsids (CCMV VLPs). We established a fitting algorithm based on a core-shell model to reckon the inner and outer diameters as well as the total mass of encapsulated RNA based on the reading of Bartlett and Ottewill [124] and Kuzmanovic *et al.* [125].

A.1 Fitting algorithm, core-shell model and Schulz distribution

We considered a solution of viral particles that we modeled with a core-shell model. We assumed that there is a distribution on the core radius r_c while the shell thickness Δ is fixed for all spheres. Bartlett and Ottewill [124] developed such a model on colloidal particles to compute the scattering intensities. The scattering profile had the approximate form

$$\rho(r) = \begin{cases} \rho_c, & r < r_c \\ \rho_s, & r_c < r < r_c + \Delta \end{cases} \tag{A.1}$$

where the spherical core of radius r_c and scattering length density ρ_c is surrounded by a shell of thickness Δ and scattering length density ρ_s. The corresponding particle scattering amplitude is expressed as

$$F(x) = \frac{4\pi}{q^3}(\rho_s - \rho_c)\{\gamma j(x + \delta x) - j(x)\} \tag{A.2}$$

where x is the dimensionless variable qr_c, δ =Δ/ρ_c, and the function $j(x)$=$\sin x - x\cos x$. ρ_m is the scattering length density of the suspension medium. γ is the scale medium contrast, $\gamma = (\rho_m - \rho_s)/(\rho_c - \rho_s)$, and determines the relative proportion of the scattering from the core or shell. Considering a distribution only of the particle core sizes while the thickness Δ is fixed for all spheres, the single particle form factor $F^2(qr_c)$ is replaced by the size average

$$\overline{F^2}(q\overline{r}_c) = \int_0^\infty G(r_c)F^2(qr_c)dr_c, \tag{A.3}$$

where $G(r_c)$ is the normalized probability of finding a particle with a core radius between r_c and $r_c + dr_c$, and $\overline{r}_c$ is the mean core radius.

Here, a Schulz distribution [126] is assumed and the normalized form of this distribution is [124]:

$$G(r_c) = \frac{r_c^Z}{\Gamma(Z+1)}\left(\frac{Z+1}{\overline{r}_c}\right)^{Z+1} \times exp\left[-\frac{r_c}{\overline{r}_c}(Z+1)\right], \tag{A.4}$$

where Γ is the Gamma function and Z is related to the polydispersity σ_c of the particle core radius by the expression :

$$\sigma_c^2 = \frac{1}{Z+1} \tag{A.5}$$

We used the analytical expression of the form factor $\overline{F^2}(q\overline{r}_c)$ given in eq.(18) from ref.[124] that we implemented in the fitting algorithm.

In the fitting algorithm, the fixed parameters were:

- the scattering length density for the coat protein : ρ_s= 12.0 $\times$ 10^{-6} Å^{-2},
- the scattering length density for the RNA : ρ_c= 16.4 $\times$ 10^{-6} Å^{-2} [125, 127],
- the scattering length density for the solvent or water : ρ_m= 9.4 $\times$ 10^{-6} Å^{-2}

The physical parameters to be determined from the fit were:

- the mean core radius $\overline{r}_c$,
- the polydispersity σ_c of the particle core radius,
- the thickness of the shell H,
- the scattering length density ρ_c of the core linked with the volume fraction of RNA in the core X by the expression

$$\rho_c = X\rho_{RNA} + (1-X)\rho_{solvent} \tag{A.6}$$

A scale factor and a constant background were also implemented as additional fit parameters.

A.2 Results

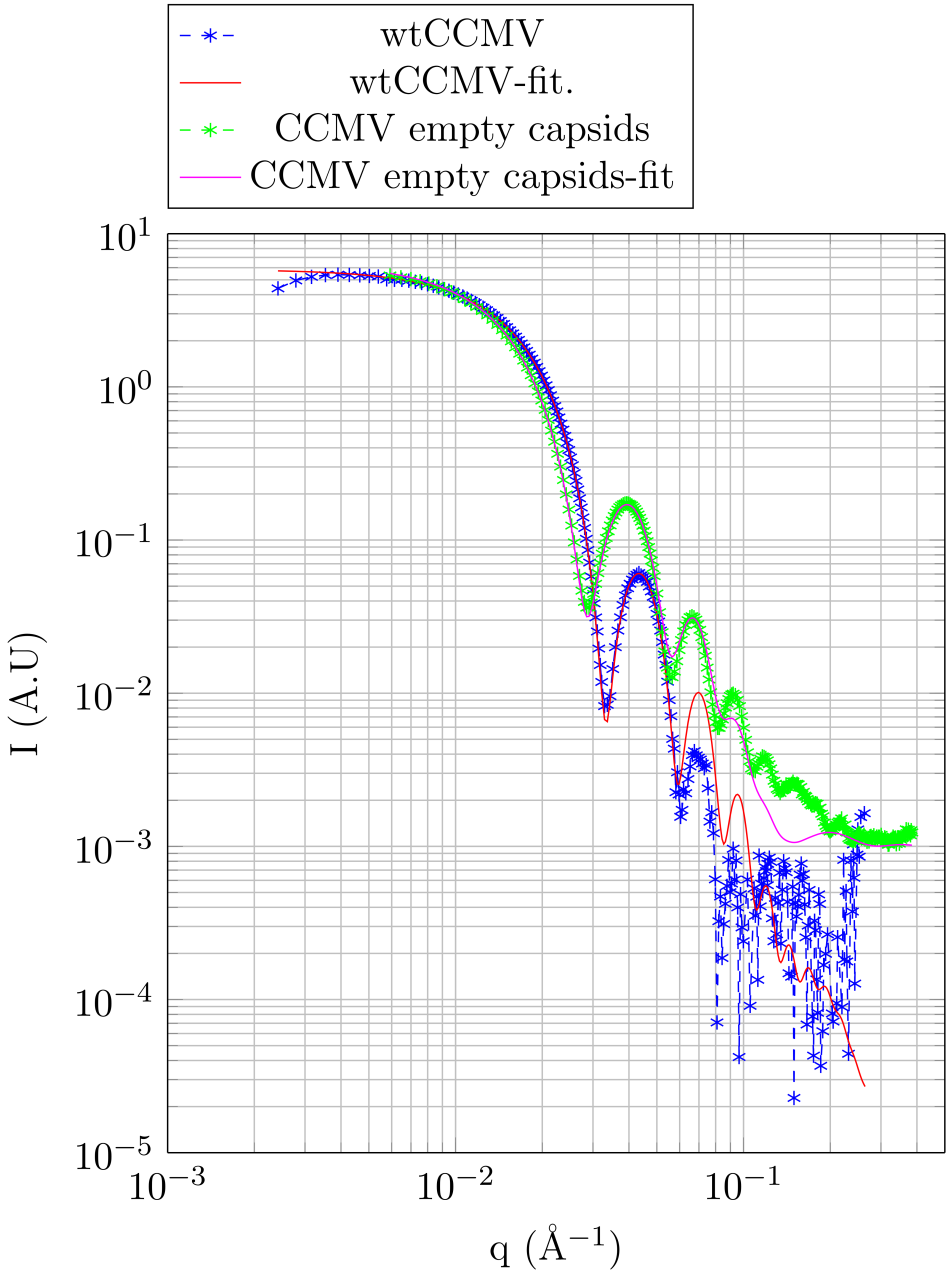


Figure A-1: Superimposed scattering spectra of wtCCMV and CCMV VLPs and results from the fit.

The results from the fit show that the dimensions of both wtCCMV and CCMV VLPs are similar. The difference comes from the presence of encapsulated RNA. X=4.5 % suggests that a small quantity of RNA was present in the solution and encapsulated by the CCMV proteins. One should bear in mind however that the

Parameters	wtCCMV	CCMV VLPs
$\overline{r}_c$ (Å)	83	88
σ_c	0.07	0.10
H (Å)	46	43
X	28 %	4.5 %

Table A.1: Quantification of RNA in wtCCMV and CCMV VLPs by SAXS : Results from the fit.

addition of the parameter X as a fit parameter could just improve the fitting.

Given the dimensions of the capsid, one can estimate the volume of the core and the volume of the RNA V_{RNA} and estimate the molar mass of encapsulated RNA :

$$M_{RNA} = \frac{V_{RNA}}{\overline{V}_{RNA}} \times N_A = 830 \text{ kDa} \tag{A.7}$$

where N_A is the Avogadro number and $\overline{V}_{RNA}$=0.50 cm$^3/g$ the partial specific volume of the ssRNA [128, 129].

This value is close to the reported 1 MDa of encapsulated RNA in wtCCMV.

Bibliography

[1] G. et al., "Cone-like virus, hiv." http://virus.chem.ucla.edu/icosahedral_symmetry.

[2] C. Foundation, "Capsid shapes in viruses." http://www.ck12.org.

[3] F. Crick and J. Watson, "Structure of small viruses.," *Nature*, vol. 177, no. 4506, pp. 473–475, 1956.

[4] R. E. Franklin and A. Klug, "The nature of the helical groove on the tobacco mosaic virus particle x-ray diffraction studies," *Biochim. Biophys. Acta*, vol. 19, pp. 403–416, 1956.

[5] J. D. Watson, "The structure of tobacco mosaic virus: I. x-ray evidence of a helical arrangement of sub-units around the longitudinal axis," *Biochim. Biophys. Acta*, vol. 13, pp. 10–19, 1954.

[6] D. Caspar and A. Klug, "Physical principles in the construction of regular viruses," in *Cold Spring Harbor symposia on quantitative biology*, vol. 27, pp. 1–24, Cold Spring Harbor Laboratory Press, 1962.

[7] J. E. Johnson and J. a. Speir, "Quasi-equivalent viruses: a paradigm for protein assemblies.," *J. Mol. Biol.*, vol. 269, pp. 665–75, June 1997.

[8] M. F. Hagan, "Modeling viral capsid assembly," *Adv. Chem. Phys.*, vol. 155, p. 1, 2014.

[9] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick, "Origin of icosahedral symmetry in viruses," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, no. 44, pp. 15556–15560, 2004.

[10] I. Rayment, T. Baker, D. Caspar, and W. Murakami, "Polyoma virus capsid structure at 22.5 åresolution," *Nature*, vol. 295, no. 5845, p. 110, 1982.

[11] R. Liddington, Y. Yan, J. Moulai, R. Sahli, T. Benjamin, and S. Harrison, "Structure of simian virus 40 at 3.8 åresolution," *Nature*, vol. 354, no. 6351, pp. 278–284, 1991.

[12] R. J. Kuhn, W. Zhang, M. G. Rossmann, S. V. Pletnev, J. Corver, E. Lenches, C. T. Jones, S. Mukhopadhyay, P. R. Chipman, E. G. Strauss, *et al.*, "Structure of dengue virus: implications for flavivirus organization, maturation, and fusion," *Cell*, vol. 108, no. 5, pp. 717–725, 2002.

[13] T. Douglas and M. Young, "Viruses: making friends with old foes," *Science*, vol. 312, no. 5775, pp. 873–875, 2006.

[14] M.-C. Daniel, I. B. Tsvetkova, Z. T. Quinkert, A. Murali, M. De, V. M. Rotello, C. C. Kao, and B. Dragnea, "Role of surface charge density in nanoparticle-templated assembly of bromovirus protein cages," *ACS nano*, vol. 4, no. 7, pp. 3853–3860, 2010.

[15] J. Sun, C. DuFort, M.-C. Daniel, A. Murali, C. Chen, K. Gopinath, B. Stein, M. De, V. M. Rotello, A. Holzenburg, *et al.*, "Core-controlled polymorphism in virus-like particles," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 104, no. 4, pp. 1354–1359, 2007.

[16] J. A. Speir, S. Munshi, G. Wang, T. S. Baker, and J. E. Johnson, "Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by x-ray crystallography and cryo-electron microscopy," *Structure*, vol. 3, no. 1, pp. 63–78, 1995.

[17] J. Tang, J. M. Johnson, K. A. Dryden, M. J. Young, A. Zlotnick, and J. E. Johnson, "The role of subunit hinges and molecular "switches" in the control of viral capsid polymorphism," *J. Struct. Biol.*, vol. 154, no. 1, pp. 59–67, 2006.

[18] J. Bancroft, G. Hills, and R. Markham, "A study of the self-assembly process in a small spherical virus formation of organized structures from protein subunits in vitro," *Virology*, vol. 31, no. 2, pp. 354–379, 1967.

[19] Q. Wang, E. Kaltgrad, T. Lin, J. E. Johnson, and M. Finn, "Natural supramolecular building blocks: wild-type cowpea mosaic virus," *Chem. Biol.*, vol. 9, no. 7, pp. 805–811, 2002.

[20] Q. Wang, T. R. Chan, R. Hilgraf, V. V. Fokin, K. B. Sharpless, and M. Finn, "Bioconjugation by copper (i)-catalyzed azide-alkyne [3+ 2] cycloaddition," *J. Am. Chem. Soc.*, vol. 125, no. 11, pp. 3192–3193, 2003.

[21] Y. J. Lee, H. Yi, W.-J. Kim, K. Kang, D. S. Yun, M. S. Strano, G. Ceder, and A. M. Belcher, "Fabricating genetically engineered high-power lithium-ion batteries using multiple virus genes," *Science*, vol. 324, no. 5930, pp. 1051–1055, 2009.

[22] P. E. Prevelige Jr, D. Thomas, and J. King, "Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells.," *Biophys. J.*, vol. 64, no. 3, p. 824, 1993.

[23] A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres, "A theoretical model successfully identifies features of hepatitis b virus capsid assembly," *Biochemistry*, vol. 38, no. 44, pp. 14644–14652, 1999.

[24] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, "Mechanism of capsid assembly for an icosahedral plant virus," *Virology*, vol. 277, no. 2, pp. 450–456, 2000.

[25] G. L. Casini, D. Graham, D. Heine, R. L. Garcea, and D. T. Wu, "In vitro papillomavirus capsid assembly analyzed by light scattering," *Virology*, vol. 325, no. 2, pp. 320–327, 2004.

[26] C. Chen, C. C. Kao, and B. Dragnea, "Self-assembly of brome mosaic virus capsids: Insights from shorter time-scale experiments†," *J. Phys. Chem. A*, vol. 112, no. 39, pp. 9405–9412, 2008.

[27] G. Tresset, C. Le Coeur, J.-F. Bryche, M. Tatou, M. Zeghal, A. Charpilienne, D. Poncet, D. Constantin, and S. Bressanelli, "Norovirus capsid proteins self-assemble through biphasic kinetics via long-lived stave-like intermediates," *J. Am. Chem. Soc.*, vol. 135, no. 41, pp. 15373–15381, 2013.

[28] C. Uetrecht, I. M. Barbu, G. K. Shoemaker, E. van Duijn, and A. J. Heck, "Interrogating viral capsid assembly with ion mobility–mass spectrometry," *Nat. Chem.*, vol. 3, no. 2, pp. 126–132, 2011.

[29] G. K. Shoemaker, E. van Duijn, S. E. Crawford, C. Uetrecht, M. Baclayon, W. H. Roos, G. J. Wuite, M. K. Estes, B. V. Prasad, and A. J. Heck, "Norwalk virus assembly and stability monitored by mass spectrometry," *Mol. Cell. Proteomics*, vol. 9, no. 8, pp. 1742–1751, 2010.

[30] W. H. Roos, R. Bruinsma, and G. J. L. Wuite, "Physical virology," *Nat. Phys.*, vol. 6, no. 10, pp. 733–743, 2010.

[31] D. Chandler, "Interfaces and the driving force of hydrophobic assembly," *Nature*, vol. 437, no. 7059, pp. 640–647, 2005.

[32] A. J. Patel, P. Varilly, S. N. Jamadagni, M. F. Hagan, D. Chandler, and S. Garde, "Sitting at the edge: How biomolecules use hydrophobicity to tune their interactions and function," *J. Phys. Chem. B*, vol. 116, no. 8, pp. 2498–2503, 2012.

[33] P. Ceres and A. Zlotnick, "Weak protein-protein interactions are sufficient to drive assembly of hepatitis b virus capsids," *Biochemistry*, vol. 41, no. 39, pp. 11525–11531, 2002.

[34] J. M. Johnson, J. Tang, Y. Nyame, D. Willits, M. J. Young, and A. Zlotnick, "Regulating self-assembly of spherical oligomers," *Nano Lett.*, vol. 5, no. 4, pp. 765–770, 2005.

[35] M. Castelnovo, D. Muriaux, and C. Faivre-Moskalenko, "Entropic control of particle sizes during viral self-assembly," *New J. Phys.*, vol. 15, no. 3, p. 035028, 2013.

[36] P. van der Schoot and R. Zandi, "Kinetic theory of virus capsid assembly," *Phys. Biol.*, vol. 4, no. 4, p. 296, 2007.

[37] D. Endres and A. Zlotnick, "Model-based analysis of assembly kinetics for virus capsids or other spherical polymers," *Biophys. J.*, vol. 83, no. 2, pp. 1217–1230, 2002.

[38] S. Kler, R. Asor, C. Li, A. Ginsburg, D. Harries, A. Oppenheim, A. Zlotnick, and U. Raviv, "Rna encapsidation by sv40-derived nanoparticles follows a rapid two-state mechanism," *J. Am. Chem. Soc.*, vol. 134, no. 21, pp. 8823–8830, 2012.

[39] R. Schwartz, P. W. Shor, P. E. Prevelige, and B. Berger, "Local rules simulation of the kinetics of virus capsid self-assembly," *Biophys. J.*, vol. 75, no. 6, pp. 2626–2636, 1998.

[40] B. Berger, P. W. Shor, L. Tucker-Kellogg, and J. King, "Local rule-based theory of virus shell assembly," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 91, no. 16, pp. 7732–7736, 1994.

[41] R. Schwartz, R. L. Garcea, and B. Berger, ""local rules" theory applied to polyomavirus polymorphic capsid assemblies," *Virology*, vol. 268, no. 2, pp. 461–470, 2000.

[42] B. Berger, J. King, R. Schwartz, and P. W. Shor, "Local rule mechanism for selecting icosahedral shell geometry," *Discrete Appl. Math.*, vol. 104, no. 1, pp. 97–111, 2000.

[43] D. Rapaport, J. Johnson, and J. Skolnick, "Supramolecular self-assembly: molecular dynamics modeling of polyhedral shell formation," *Comput. Phys. Commun.*, vol. 121, pp. 231–235, 1999.

[44] D. Rapaport, "Self-assembly of polyhedral shells: a molecular dynamics study," *Phys. Rev. E*, vol. 70, no. 5, p. 051905, 2004.

[45] D. Rapaport, "Molecular dynamics simulation of reversibly self-assembling shells in solution using trapezoidal particles," *Phys. Rev. E*, vol. 86, no. 5, p. 051917, 2012.

[46] H. D. Nguyen, V. S. Reddy, and C. L. Brooks, "Deciphering the kinetic mechanism of spontaneous self-assembly of icosahedral capsids," *Nano Lett.*, vol. 7, no. 2, pp. 338–344, 2007.

[47] R. F. Allison, M. Janda, and P. Ahlquist, "Infectious in vitro transcripts from cowpea chlorotic mottle virus cdna clones and exchange of individual rna components with brome mosaic virus.," *J. Virol.*, vol. 62, no. 10, pp. 3581–3588, 1988.

[48] A. M. Dzianott and J. I. Bujarski, "The nucleotide sequence and genome organization of the rna-1 segment in two bromoviruses: broad bean mottle virus and cowpea chlorotic mottle virus," *Virology*, vol. 185, no. 2, pp. 553–562, 1991.

[49] J. Chidlow and J. Tremaine, "Limited hydrolysis of cowpea chlorotic mottle virus by trypsin and chymotrypsin," *Virology*, vol. 43, no. 1, pp. 267–278, 1971.

[50] G. Vriend, M. Hemminga, B. Verduin, J. De Wit, and T. Schaafsma, "Segmental mobility involved in protein—rna interaction in cowpea chlorotic mottle virus," *FEBS Lett.*, vol. 134, no. 2, pp. 167–171, 1981.

[51] G. Vriend, B. Verduin, and M. Hemminga, "Role of the n-terminal part of the coat protein in the assembly of cowpea chlorotic mottle virus: A 500 mhz proton nuclear magnetic resonance study and structural calculations," *J. Mol. Biol.*, vol. 191, no. 3, pp. 453–460, 1986.

[52] K. W. Adolph and P. Butler, "Studies on the assembly of a spherical plant virus: Iii. reassembly of infectious virus under mild conditions," *J. Mol. Biol.*, vol. 109, no. 2, pp. 345–357, 1977.

[53] X. Zhao, J. M. Fox, N. H. Olson, T. S. Baker, and M. J. Young, "In vitro assembly of cowpea chlorotic mottle virus from coat protein expressed in escherichia coli and in vitro-transcribed viral cdna," *Virology*, vol. 207, no. 2, pp. 486–494, 1995.

[54] A. Ali and M. J. Roossinck, "Rapid and efficient purification of cowpea chlorotic mottle virus by sucrose cushion ultracentrifugation," *J. Virol. Methods*, vol. 141, no. 1, pp. 84–86, 2007.

[55] M. C. Aragones, *The cowpea chlorotic mottle virus as a building block in nanotechnology*. PhD thesis, Radboud University Nijmegen, Netherlands, 2010.

[56] J.-P. Michel, M. Gingery, and L. Lavelle, "Efficient purification of bromoviruses by ultrafiltration," *J. Virol. Methods*, vol. 122, no. 2, pp. 195–198, 2004.

[57] E. Hiebert, J. Bancroft, and C. Bracker, "The assembly in vitro of some small spherical viruses, hybrid viruses, and other nucleoproteins," *Virology*, vol. 34, no. 3, pp. 492–508, 1968.

[58] L. Lavelle, J.-P. Michel, and M. Gingery, "The disassembly, reassembly and stability of ccmv protein capsids," *J. Virol. Methods*, vol. 146, no. 1, pp. 311–316, 2007.

[59] L. Lavelle, M. Gingery, M. Phillips, W. Gelbart, C. Knobler, R. Cadena-Nava, J. Vega-Acosta, L. Pinedo-Torres, and J. Ruiz-Garcia, "Phase diagram of self-assembled viral capsid protein polymorphs," *J. Phys. Chem. B*, vol. 113, no. 12, pp. 3813–3819, 2009.

[60] B. Verduin, "The preparation of ccmv-protein in connection with its association into a spherical particle," *FEBS Lett.*, vol. 45, no. 1, pp. 50–54, 1974.

[61] K. Adolph and P. Butler, "Studies on the assembly of a spherical plant virus: I. states of aggregation of the isolated protein," *J. Mol. Biol.*, vol. 88, no. 2, pp. 327–341, 1974.

[62] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, "Viral self-assembly as a thermodynamic process," *Phys. Rev. Lett.*, vol. 90, no. 24, p. 248101, 2003.

[63] R. D. Cadena-Nava, Y. Hu, R. F. Garmann, B. Ng, A. N. Zelikin, C. M. Knobler, and W. M. Gelbart, "Exploiting fluorescent polymers to probe the self-assembly of virus-like particles," *J. Phys. Chem. B*, vol. 115, no. 10, pp. 2386–2391, 2011.

[64] T. Douglas and M. Young, "Host–guest encapsulation of materials by assembled virus protein cages," *Nature*, vol. 393, no. 6681, pp. 152–155, 1998.

[65] C. B. Chang, C. M. Knobler, W. M. Gelbart, and T. G. Mason, "Curvature dependence of viral protein structures on encapsidated nanoemulsion droplets," *ACS Nano*, vol. 2, no. 2, pp. 281–286, 2008.

[66] A. de la Escosura, J. Roeland, J. Jeroen, *et al.*, "Viral capsids as templates for the production of monodisperse prussian blue nanoparticles," *Chem. Commun.*, no. 13, pp. 1542–1544, 2008.

[67] E. Gillitzer, D. Willits, M. Young, and T. Douglas, "Chemical modification of a viral cage for multivalent presentation," *Chem. Commun.*, no. 20, pp. 2390–2391, 2002.

[68] P. A. Suci, D. L. Berglund, L. Liepold, S. Brumfield, B. Pitts, W. Davison, L. Oltrogge, K. O. Hoyt, S. Codd, P. S. Stewart, *et al.*, "High-density targeting of a viral multifunctional nanoplatform to a pathogenic, biofilm-forming bacterium," *Chem. Biol.*, vol. 14, no. 4, pp. 387–398, 2007.

[69] P. A. Suci, Z. Varpness, E. Gillitzer, T. Douglas, and M. Young, "Targeting and photodynamic killing of a microbial pathogen using protein cage architectures functionalized with a photosensitizer," *Langmuir*, vol. 23, no. 24, pp. 12280–12286, 2007.

[70] L. Feigin, D. I. Svergun, and G. W. Taylor, *Structure analysis by small-angle X-ray and neutron scattering*. Springer, 1987.

[71] D. I. Svergun and M. H. Koch, "Small-angle scattering studies of biological macromolecules in solution," *Rep. Prog. Phys.*, vol. 66, no. 10, p. 1735, 2003.

[72] T. Zemb and P. Lindner, *Neutrons, X-rays and light: scattering methods applied to soft condensed matter*. North-Holland, 2002.

[73] O. Spalla, "General theorems in small-angle scattering," in *Neutrons, X-rays and light: scattering methods applied to soft condensed matter*, pp. 54–58, North-Holland, 2002.

[74] A. Guinier and G. Fournet, *Small-angle scattering of X-rays*. Structure of matter series, Wiley, 1955.

[75] P. Panine, S. Finet, T. Weiss, and T. Narayanan, "Probing fast kinetics in complex fluids by combined rapid mixing and small-angle x-ray scattering," *Adv. Colloid Interface Sci.*, vol. 127, no. 1, pp. 9–18, 2006.

[76] C. A. Dreiss, K. S. Jack, and A. P. Parker, "On the absolute calibration of bench-top small-angle x-ray scattering instruments: a comparison of different standard methods," *J. Appl. Crystallogr.*, vol. 39, no. 1, pp. 32–38, 2006.

[77] B. R. Pauw, ""the impossible project" - how to do a perfect saxs measurement, 2011." http://www.lookingatnothing.com/wp-content/uploads/2011/04/imp_userguide.pdf.

[78] E. R. Henry and J. Hofrichter, "Singular value decomposition: application to analysis of experimental data," *Methods Enzymol.*, vol. 210, pp. 129–192, 1992.

[79] L. Zimányi, Á. Kulcsár, J. K. Lanyi, D. F. Sears, and J. Saltiel, "Singular value decomposition with self-modeling applied to determine bacteriorhodopsin intermediate spectra: analysis of simulated data," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, no. 8, pp. 4408–4413, 1999.

[80] L. Zimányi, Á. Kulcsár, J. K. Lanyi, D. F. Sears, and J. Saltiel, "Intermediate spectra and photocycle kinetics of the asp96-asn mutant bacteriorhodopsin determined by singular value decomposition with self-modeling," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 96, no. 8, pp. 4414–4419, 1999.

[81] S. Doniach, "Changes in biomolecular conformation seen by small angle x-ray scattering," *Chem. Rev.*, vol. 101, no. 6, pp. 1763–1778, 2001.

[82] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.

[83] R. H. Byrd, M. E. Hribar, and J. Nocedal, "An interior point algorithm for large-scale nonlinear programming," *SIAM J. Optim.*, vol. 9, no. 4, pp. 877–900, 1999.

[84] R. H. Byrd, J. C. Gilbert, and J. Nocedal, "A trust region method based on interior point techniques for nonlinear programming," *Math. Prog.*, vol. 89, no. 1, pp. 149–185, 2000.

[85] L. F. Shampine and M. W. Reichelt, "The matlab ode suite," *SIAM J. Sci. Comput.*, vol. 18, no. 1, pp. 1–22, 1997.

[86] L. F. Shampine, M. W. Reichelt, and J. A. Kierzenka, "Solving index-1 daes in matlab and simulink," *SIAM Rev.*, vol. 41, no. 3, pp. 538–552, 1999.

[87] M. V. Petoukhov, P. V. Konarev, A. G. Kikhney, and D. I. Svergun, "Atsas 2.1-towards automated and web-supported small-angle scattering data analysis," *J. Appl. Crystallogr.*, 2007.

[88] B. D. Wilts, I. A. Schaap, and C. F. Schmidt, "Swelling and softening of the cowpea chlorotic mottle virus in response to ph shifts," *Biophys. J.*, vol. 108, no. 10, pp. 2541–2549, 2015.

[89] D. Law-Hine, A. K. Sahoo, V. Bailleux, M. Zeghal, S. Prevost, P. K. Maiti, S. Bressanelli, D. Constantin, and G. Tresset, "Reconstruction of the disassembly pathway of an icosahedral viral capsid and shape determination of two successive intermediates," *J. Phys. Chem. Lett.*, vol. 6, no. 17, pp. 3471–3476, 2015.

[90] D. I. Svergun, C. Barberato, and M. H. J. Koch, "Crysol - a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates.," *J. Appl. Crystallogr.*, vol. 28, pp. 768–773, 1995.

[91] D. Svergun, "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing," *Biophys. J.*, vol. 76, no. 6, pp. 2879–2886, 1999.

[92] V. V. Volkov and D. I. Svergun, "Uniqueness of ab initio shape determination in small-angle scattering," *J. Appl. Crystallogr.*, vol. 36, no. 3, pp. 860–864, 2003.

[93] V. S. Reddy, H. A. Giesing, R. T. Morton, A. Kumar, C. B. Post, C. L. Brooks, and J. E. Johnson, "Energetics of quasiequivalence: computational analysis of protein-protein interactions in icosahedral viruses," *Biophys. J.*, vol. 74, no. 1, pp. 546–558, 1998.

[94] V. S. Reddy and J. E. Johnson, "Structure-derived insights into virus assembly," *Adv. Virus Res.*, vol. 64, pp. 45–68, 2005.

[95] D. Svergun, "Solution scattering from biopolymers: advanced contrast-variation data analysis," *Acta Crystallogr. Sect. A*, vol. 50, no. 3, pp. 391–402, 1994.

[96] D. Endres, M. Miyahara, P. Moisant, and A. Zlotnick, "A reaction landscape identifies the intermediates critical for self-assembly of virus capsids and other polyhedral structures," *Protein Sci.*, vol. 14, no. 6, pp. 1518–1525, 2005.

[97] C. D. Putnam, M. Hammel, G. L. Hura, and J. A. Tainer, "X-ray solution scattering (saxs) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution," *Q. Rev. Biophys.*, vol. 40, no. 03, pp. 191–285, 2007.

[98] G. Polles, G. Indelicato, R. Potestio, P. Cermelli, R. Twarock, and C. Micheletti, "Mechanical and assembly units of viral capsids identified via quasi-rigid domain decomposition," *PLoS Comput. Biol.*, vol. 9, p. e1003331, 11 2013.

[99] R. Tuma, H. Tsuruta, K. H. French, and P. E. Prevelige, "Detection of intermediates and kinetic control during assembly of bacteriophage p22 procapsid," *J. Mol. Biol.*, vol. 381, no. 5, pp. 1395–1406, 2008.

[100] H. Fraenkel-Conrat and R. C. Williams, "Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 41, no. 10, p. 690, 1955.

[101] R. D. Cadena-Nava, M. Comas-Garcia, R. F. Garmann, A. Rao, C. M. Knobler, and W. M. Gelbart, "Self-assembly of viral capsid protein and rna molecules of different sizes: requirement for a specific high protein/rna mass ratio," *J. Virol.*, vol. 86, no. 6, pp. 3318–3326, 2012.

[102] M. Comas-Garcia, R. D. Cadena-Nava, A. Rao, C. M. Knobler, and W. M. Gelbart, "In vitro quantification of the relative packaging efficiencies of single-stranded rna molecules by viral capsid protein," *J. Virol.*, vol. 86, no. 22, pp. 12271–12282, 2012.

[103] R. F. Garmann, M. Comas-Garcia, A. Gopal, C. M. Knobler, and W. M. Gelbart, "The assembly pathway of an icosahedral single-stranded rna virus depends on the strength of inter-subunit attractions," *J. Mol. Biol.*, vol. 426, no. 5, pp. 1050–1060, 2014.

[104] J. M. Johnson, D. A. Willits, M. J. Young, and A. Zlotnick, "Interaction with capsid protein alters rna structure and the pathway for in vitro assembly of cowpea chlorotic mottle virus," *J. Mol. Biol.*, vol. 335, no. 2, pp. 455–464, 2004.

[105] O. M. Elrad and M. F. Hagan, "Mechanisms of size control and polymorphism in viral capsid assembly," *Nano Lett.*, vol. 8, no. 11, pp. 3850–3857, 2008.

[106] O. M. Elrad and M. F. Hagan, "Encapsulation of a polymer by an icosahedral virus," *Phys. Biol.*, vol. 7, no. 4, p. 045003, 2010.

[107] A. McPherson, "Micelle formation and crystallization as paradigms for virus assembly," *Bioessays*, vol. 27, no. 4, pp. 447–458, 2005.

[108] J. P. Mahalik and M. Muthukumar, "Langevin dynamics simulation of polymer-assisted virus-like assembly," *J. Chem. Phys.*, vol. 136, no. 13, 2012.

[109] T. Hu, R. Zhang, and B. Shklovskii, "Electrostatic theory of viral self-assembly," *Phys. A Stat. Mech. its Appl.*, vol. 387, pp. 3059–3064, May 2008.

[110] T. Hu and B. Shklovskii, "Kinetics of viral self-assembly: Role of the single-stranded RNA antenna," *Phys. Rev. E*, vol. 75, p. 051901, May 2007.

[111] R. Zandi and P. Van der Schoot, "Size regulation of ss-rna viruses," *Biophys. J.*, vol. 96, no. 1, pp. 9–20, 2009.

[112] A. S. Petrov and S. C. Harvey, "Packaging double-helical dna into viral capsids: structures, forces, and energetics," *Biophys. J.*, vol. 95, no. 2, pp. 497–502, 2008.

[113] M. S. Dhason, J. C.-Y. Wang, M. F. Hagan, and A. Zlotnick, "Differential assembly of hepatitis b virus core protein on single-and double-stranded nucleic acid suggest the dsdna-filled core is spring-loaded," *Virology*, vol. 430, no. 1, pp. 20–29, 2012.

[114] S. Tzlil, J. T. Kindt, W. M. Gelbart, and A. Ben-Shaul, "Forces and pressures in dna packaging and release from viral capsids," *Biophys. J.*, vol. 84, no. 3, pp. 1616–1627, 2003.

[115] S. C. Kowalczykowski, L. S. Paul, N. Lonberg, J. W. Newport, J. A. McSwiggen, and P. H. Von Hippel, "Cooperative and noncooperative binding of protein ligands to nucleic acid lattices: experimental approaches to the determination of thermodynamic parameters," *Biochemistry*, vol. 25, no. 6, pp. 1226–1240, 1986.

[116] J. D. McGhee and P. H. von Hippel, "Theoretical aspects of dna-protein interactions: co-operative and non-co-operative binding of large ligands to a one-dimensional homogeneous lattice," *J. Mol. Biol.*, vol. 86, no. 2, pp. 469–489, 1974.

[117] A. Zlotnick, "To build a virus capsid: an equilibrium model of the self assembly of polyhedral protein complexes," *J. Mol. Biol.*, vol. 241, no. 1, pp. 59–67, 1994.

[118] A. Zlotnick, J. Z. Porterfield, and J. C.-Y. Wang, "To build a virus on a nucleic acid substrate," *Biophys. J.*, vol. 104, no. 7, pp. 1595–1604, 2013.

[119] A. Kobayashi and Y. Ehara, "In vitro encapsidation of cucumber mosaic virus rna species," *Ann. Phytopathol. Soc. Jpn*, vol. 61, no. 2, pp. 99–102, 1995.

[120] Y. Hu, R. Zandi, A. Anavitarte, C. M. Knobler, and W. M. Gelbart, "Packaging of a polymer by a viral capsid: the interplay between polymer length and capsid size.," *Biophys. J.*, vol. 94, pp. 1428–36, Feb. 2008.

[121] A. Kivenson and M. F. Hagan, "Mechanisms of capsid assembly around a polymer," *Biophys. J.*, vol. 99, no. 2, pp. 619–628, 2010.

[122] F. Tama and C. L. Brooks, "The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus.," *J. Mol. Biol.*, vol. 318, pp. 733–47, May 2002.

[123] H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, and L. Pollack, "Ionic strength-dependent persistence lengths of single-stranded rna and dna," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 109, no. 3, pp. 799–804, 2012.

[124] P. Bartlett and R. Ottewill, "A neutron scattering study of the structure of a bimodal colloidal crystal," *J. Chem. Phys*, vol. 96, no. 4, pp. 3306–3318, 1992.

[125] D. A. Kuzmanovic, I. Elashvili, C. Wick, C. O'Connell, and S. Krueger, "Quantification of rna in bacteriophage ms2-like viruses in solution by small-angle x-ray scattering," *Radiat. Phys. Chem.*, vol. 75, no. 3, pp. 359–368, 2006.

[126] B. H. Zimm, "The scattering of light and the radial distribution function of high polymer solutions," *J. Chem. Phys*, vol. 16, no. 12, pp. 1093–1099, 1948.

[127] D. Orthaber, A. Bergmann, and O. Glatter, "Saxs experiments on absolute scale with kratky systems using water as a secondary standard," *J. Appl. Crystallogr.*, vol. 33, no. 2, pp. 218–225, 2000.

[128] S. J. Greive, A. F. Lins, and P. H. von Hippel, "Assembly of an rna-protein complex binding of nusb and nuse (s10) proteins to boxa rna nucleates the formation of the antitermination complex involved in controlling rrna transcription in escherichia coli," *J. Biol. Chem.*, vol. 280, no. 43, pp. 36397–36408, 2005.

[129] G. Cohen and H. Eisenberg, "Deoxyribonueleate solutions: Sedimentation in a density gradient, partial specific volumes, density and refractive index increments, and preferential interactions," *Biopolymers*, vol. 6, no. 8, pp. 1077–1100, 1968.

Résumé de thèse

en français

Contents

1 Introduction générale

Les virus sont des entités biologiques connus pour leur propriétés d'interaction spécifique avec un récepteur cellulaire et d'auto-réplication. Ces entités sont incapables de se reproduire seules et ont besoin d'une cellule hôte pour le faire. Une fois que la cellule hôte spécifique est infectée, les virus se répliquent en utilisant le métabolisme et les constituants de cette cellule.

De point de vue de la physique, les virus sont perçus comme des nanoparticules qui possèdent des propriétés remarquables. Leur architecture parfaitement structurée font d'eux des systèmes particulièrement fascinants à étudier. Il est aussi frappant que l'auto-assemblage viral conduit souvent à des objets de taille unique et que ces objets maintiennent leur activité infectieuse. Les clés pour comprendre ce haut degré de perfection résident dans la compréhension des mécanismes physiques derrière l'auto-organisation des sous-unités protéiques et l'encapsulation du génome. Beaucoup d'études ont été faites et ces mécanismes ne sont pas encore entièrement compris.

Les virus ont longtemps été considérés comme des agents infectieux à l'origine de maladies sérieuses et, par conséquent, systématiquement dangereux. Mais depuis peu leur étonnantes propriétés physiques ont été utilisées à des fins plus positives. Dans le contexte de la thérapie génique par exemple, les virus sont perçus comme de potentiels transporteurs d'agents thérapeutiques à cause de leur propriété d'interaction spécifique. Du fait de leur propriétés remarquables de réplication et d'auto-assemblage, plusieurs applications ont été trouvées dans le domaine des nanomatériaux. Les protéines virales sont maintenant utilisées comme briques fondamentales pour fabriquer des assemblages supramoléculaires complexes.

Les études expérimentales de physique sur les virus ont souvent été réalisées à l'équilibre thermodynamique. La dynamique d'auto-assemblage, cependant, est plus délicate à étudier. D'un point de vue théorique, le processus d'auto-assemblage a été étudié par le biais de simulations numériques, plus particulièrement des simulations de dynamique moléculaire. Expérimentalement, l'auto-assemblage ne peut être analysé qu'à partir des informations extraites lors de la cinétique.

Cette thèse s'inscrit dans le contexte spécifique de l'étude de cinétique d'auto-assemblage viral. La première partie de ce résumé introduira le système viral étudié pendant cette thèse, le virus de la marbrure chlorotique de la cornille (CCMV). La technique de diffusion des rayons X aux petits angles résolue en temps (TR-SAXS), très utilisée dans cette thèse, sera également présentée. Les aspects mathématiques sur l'analyse de données ainsi que les modèles numériques implémentés seront introduits. Une étude sur le désassemblage de capsides vides de CCMV sera entièrement exposée dans une deuxième partie. On verra notamment que la technique TR-SAXS permet d'extraire à la fois de l'information structurale et temporelle. Les résultats de cette étude ont été publiés en Aout 2015. Une troisième partie se consacrera à l'étude de l'auto-assemblage de capsides vides de CCMV. Cette partie illustrera la complexité d'une telle étude et la méthode adaptée choisie pour parvenir à décrire la cinétique de manière quantitative. Enfin, la dernière partie explorera la cinétique d'encapsulation d'ARN simple brin. Ce projet est toujours en cours d'étude mais les principales données analysées seront examinées brièvement.

2 Présentation du système viral (CCMV) et de la technique TR-SAXS

2.1 Virus de la marbrure chlorotique de la cornille - Cowpea Chlorotic Mottle Virus (CCMV)

Le virus de la marbrure chlorotique de la cornille (CCMV) est un virus de plante qui infecte la plante cornille. Les termes "marbrure" et "chlorotique" viennent du fait que les feuilles infectées développent des taches jaunâtres. Le CCMV appartient à la famille des *Bromovidae*. Son génome est composé de 4 types d'ARN simple brin: ARN1 (3,171 nucleotides (nt)); ARN2, 2,774 nt; ARN3, 2,173 nt; et, ARN4, 824 nt. La structure cristallographique du virus natif à été résolue [1] (voir Fig. 1) et l'information concernant la séquence proteique est contenue dans un fichier .pdb (reference 1ZA7).

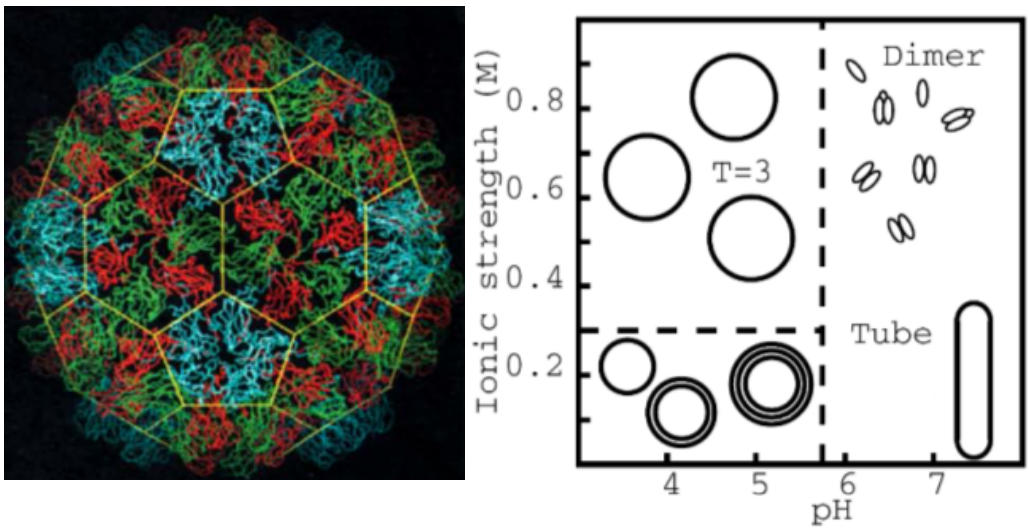


Figure 1: (*Gauche*) Structure du CCMV. Les 60 protéines des axes pentamériques sont présentées en bleu (protéines A). Les protéines B et C forment les sous-unités hexameriques et sont respectivement présentées en rouge et vert. Source: [1]. (*Droite*) Diagramme de phase des protéines de la capside du CCMV adapté de [2, 3].

La capside est composée de 180 protéines chimiquement identiques. Elle présente une symétrie icosaédrique de type T=3 selon la classification de Caspar et Klug [4]. Sa masse est de 180×20 kDa=3.6 MDa. Son diamètre est de 280 Å et l'épaisseur de la capside est de 50 Å. Les protéines sont arrangées en capsomères qui sont des pentamères (au nombre de 12) et des hexamères (20) composées de protéines notées conventionnellement protéines A,B et C. Les protéines A forment les pentamères de la capside et les protéines B et C, les hexamères (voir Fig. 1). Le N-terminal joue une rôle important dans l'encapsulation de l'ARN [5, 6] puisque ce dernier est chargé

négativement (le N-terminal possède 10 résidus cationiques). En solution, lorsque le virus est complètement désassemblé, les protéines existent sous forme de dimères [7, 8] (cf Fig. 1). Ainsi, la formation d'une capside peut être vue comme un assemblage de 60 dimères A-B et 30 dimères C-C.

2.2 Diffusion des rayons X et neutrons aux petits angles résolue en temps (TR-SAXS)

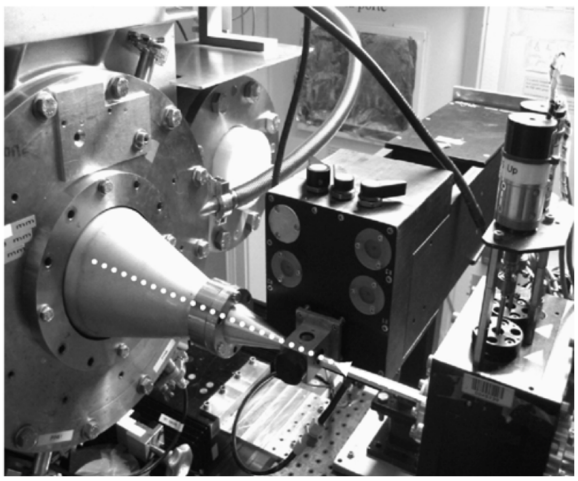


Figure 2: Illustration du dispositif experimental comprenant le stopped-flow sur la ligne ID02 (ESRF). Les flèches indiquent la direction du rayon incident[9].

Les dernières décennies ont vu le développement rapide de techniques synchrotrons pour des mesures en temps réel de cinétique rapide. Les techniques de diffusion à petits angles résolue en temps des rayons X et neutrons (TR-SAXS/SANS) sont devenues des techniques majeures pour de tels études de cinétique. Les mesures de cinétique en temps réel réfèrent souvent à l'étude de systèmes en solution sujets à une perturbation (saut de pH, dilution, etc.). La perturbation induit un changement d'état du système qui est observé en fonction du temps. Les premières étapes de la cinétique sont souvent difficiles à observer. Pour des temps très courts (typiquement une dizaine de millisecondes), il est crucial de déclencher l'acquisition avec un système automatisé qui permet de coordonner les différents appareils de mesures. Les techniques TR-SAXS/SANS requièrent surtout un appareil de haut de gamme qui permette d'effectuer une perturbation de manière rapide. L'appareil de stopped-flow est aujourd'hui indispensable pour ce genre d'études. Il permet en effet de mélanger efficacement plusieurs solutions de manière très rapide (typiquement 5 ms). Il est couplé à des détecteurs de haute performance et la haute brillance des faisceaux de rayons X et neutrons (cf Fig. 2).

2.3 Outils mathématiques pour traitement des données de TR-SAXS/SANS

2.3.1 Décomposition en Valeurs Singulières (SVD)

L'analyse des données de TR-SAXS consiste essentiellement en des fits de courbes SAXS par un modèle physique. Le premier outil mathématique dont nous nous servons est la décomposition en valeurs singulières (SVD) qui est une technique de réduction des données. Les données de TR-SAXS sont les intensités diffusées (moyennées) qui peuvent se mettre sous la forme de matrice $\mathbf{I}$. La décomposition en valeurs singulières (SVD) consiste à exprimer cette matrice sous la forme $\mathbf{I} = \mathbf{USV^T}$ où $\mathbf{U}$ est une matrice de taille $M \times N$, $\mathbf{V}$ une matrice de taille $N \times N$, est $\mathbf{S}$ une $N \times N$ matrice diagonale d'éléments positifs s_i appelés valeurs singulières de $\mathbf{I}$. Connaissant les incertitudes σ, un premier critère de réduction de données consiste à décrire les intensités avec une matrice tronquée de rank k:

$$\| \mathbf{I} - \tilde{\mathbf{I}} \|^2 = \sum_{i=k+1}^{N} s_i^2 \leq \mu\nu\sigma^2 \tag{1}$$

où σ^2 est la variance. Les paramètres μ et ν peuvent être choisis comme $\mu = M - k$ et $\nu = N - k$ [10].

Ce critère permet d'évaluer le nombre de composants qui contribuent au signal. Un deuxième critère requis consiste à considérer les facteurs d'autocorrelations C des vecteurs propres $\mathbf{U_i}$ et $\mathbf{V_i}$ définis par:

$$\begin{aligned} C(\mathbf{U_i}) &= \sum_{j=1}^{M-1} U_{j,i} U_{j+1,i} \\ C(\mathbf{V_i}) &= \sum_{j=1}^{N-1} V_{j,i} V_{j+1,i} \end{aligned} \tag{2}$$

Les vecteurs propres étant normalisés, les facteurs d'autocorrélations qui sont proches de 1 indiquent un bon rapport signal sur bruit. Les vecteurs propres correspondant peuvent être utilisés par la suite. Les facteurs d'autocorrélations qui sont proches de 0 indiquent en revanche un faible rapport signal sur bruit. Pour plus de détails, le lecteur est invité à lire l'article sur lequel se base cette section en [10].

2.3.2 Global fitting

Le terme "global fitting" réfère généralement au fit simultané d'une série de courbes. Dans nos études, les données de TR-SAXS sont des données de cinétique, où l'on collecte des spectres SAXS à différents instants de la cinétique. Les paramètres de fit correspondent essentiellement aux constantes de réaction de la cinétique, les concentrations initiales ainsi que les coefficients stoechiométriques et ordres de réaction.

Typiquement, dans une solution où coexistent L différents oligomères d'une même

protéine, l'intensité totale diffusée peut s'exprimer sous forme matricielle:

$$\begin{aligned}\mathbf{I} &= K(\overline{\nu},\Delta\rho)\begin{pmatrix} M_1^2P_1(q_1) & \cdots & M_L^2P_L(q_1) \\ \vdots & \vdots & \vdots \\ M_1^2P_1(q_M) & \cdots & M_L^2P_L(q_M)\end{pmatrix}\begin{pmatrix} c_1(t_1) & \cdots & c_1(t_N) \\ \vdots & \vdots & \vdots \\ c_L(t_1) & \cdots & c_L(t_N)\end{pmatrix} \\ &= K(\overline{\nu},\Delta\rho)\mathbf{P'c}\end{aligned} \tag{3}$$

avec $K(\overline{\nu},\Delta\rho)$ une constante qui dépend du volume spécifique $\overline{\nu}$ et du contraste de densité de longueur de diffusion $\Delta\rho$. M_l, P_l et c_l sont respectivement les masses moléculaires, facteurs de formes et concentration molaires des espèces l. Les facteurs de forme $P_l(q)$ contiennent l'information liée à la structure des espèces l alors que les facteurs $c_l(t)$ donnent l'évolution temporelle des concentrations de ces espèces.

Le nombre L d'espèces est choisi en fonction du nombre de composants qui contribuent au signal diffusé déterminé par la SVD (voir paragraphe précedent).

La procedure de global fitting consiste à trouver les facteurs de forme P_l et les paramètres $\{\xi_i\}$ de cinétique (liés à l'évolution des concentrations $\mathbf{c}(t,\{\xi_i\})$), de telle sorte à ce que la relation $\mathbf{I} \approx \mathbf{I^{fit}} = K\mathbf{P'c}(\{\xi_i\})$ soit optimale. Cela peut s'interpréter par une approche des moindres carrés qui veut que le résidu

$$\chi^2 = \frac{1}{MN-1}\sum_{m,n=1}^{M,N}\left(\frac{I_{mn}-I_{mn}^{fit}}{\sigma_{mn}}\right)^2 \tag{4}$$

soit minimum par rapport à $\mathbf{P'}$ et $\{\xi_i\}$, σ_{mn} étant les incertitudes experimentales. $\mathbf{P'}$ et $\{\xi_i\}$ sont les paramètres de fit. Le global fitting est donc un problème d'optimisation non-linéaire.

Les paramètres de cinétique $\{\xi_i\}$ gouvernent l'évolution des concentrations des espèces à travers un jeu d'équations différentielles:

$$\begin{cases} \dfrac{dc_1}{dt} &= f_1\left(c_1,\cdots,c_L;\{\xi_i\}\right) \\ \vdots & \vdots \\ \dfrac{dc_L}{dt} &= f_L\left(c_1,\cdots,c_L;\{\xi_i\}\right) \end{cases} \tag{5}$$

Les fonctions f_i dependent du modèle de cinétique utilisé. Les paramètres ξ_i incluent les coefficients stoechiometriques, les constantes de réaction et les ordres partiels de réaction.

3 Etude de la cinétique de désassemblage de capsides vides de CCMV par TR-SAXS

Nous étudions dans un premier temps la cinétique de désassemblage de capsides vides de CCMV avec la technique de TR-SAXS. Etant donné le diagramme de phase en Fig 1, à une salinité de 0.5 M, nous utilisons le stopped-flow pour provoquer un saut de pH (de pH 4.8 à pH 7.5) qui va nous permettre d'observer la cinétique de désassemblage des capsides vides en dimères. Le but de cette étude est de détecter et d'identifier les intermédiaires de réaction produits lors de la cinétique. Il s'agit également d'identifier les mécanismes réactionnels qui gouvernent cette cinétique. La résolution de la structure de ces intermédiaires et de ces mécanismes réactionnels est une autre façon de mieux comprendre les interactions en jeu entre protéines.

3.1 Approximation de Guinier & SVD

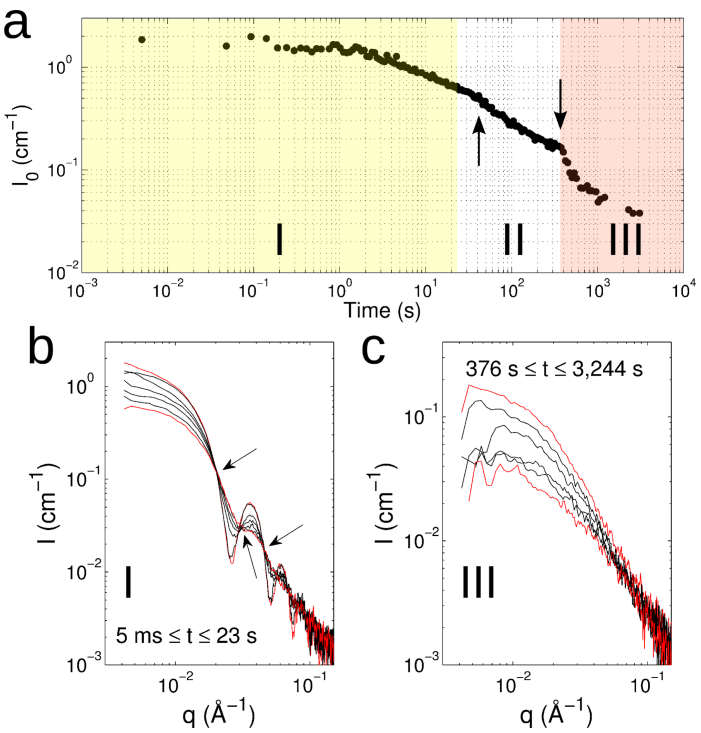


Figure 3: Données de TR-SAXS. (a) Evolution de l'intensité diffusée I_0 de 5 ms à 3,244 s. Les phases numerotées I, II and III délimitent les trois phases du désassemblage. Les flêches indiquent des changements notables de l'intensité I_0(t). (b,c) Superposition des spectres aux temps courts (phase I) et temps longs (phase III). Les spectres en rouge sont les premiers et derniers spectres de chaque phase.[11]

L'approximation de Guinier établie sur nos données de TR-SAXS nous permettent dans un premier temps de distinguer trois phases caractérisées par des changements notables de l'intensité diffusée I_0 (voir Fig. 3).

Une analyse SVD effectuée sur la totalité de la cinétique nous indique que quatre composantes ressortent du signal SAXS. Une analyse plus approfondie sur les phases I et III révèlent que deux composantes seulement contribuent au signal de l'intensité diffusée dans chacune de ces deux phases. Pour la phase I, cela peut physiquement s'interpréter par le fait que les capsides (composante 1) se dissocient en un gros intermédiaire (composante 2). Pour la phase III, il est physiquement probable que ce soit un petit intermédiaire (composante 3) qui se dissocie en dimères (composante 4).

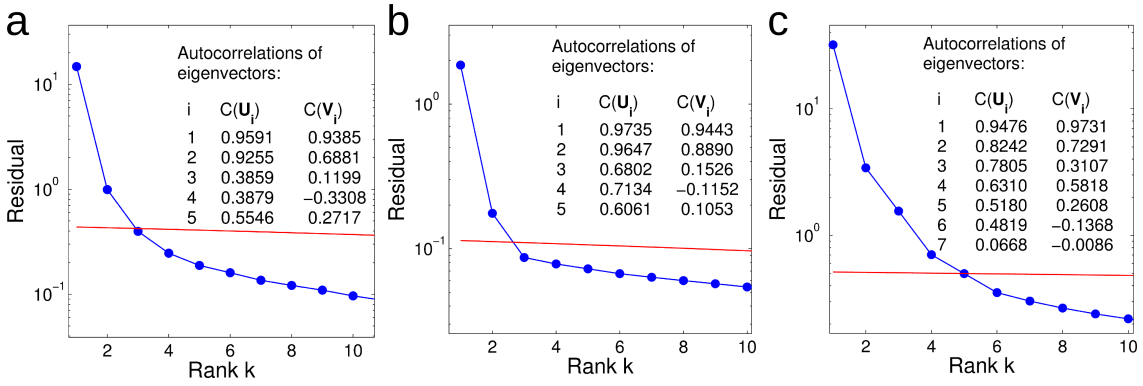


Figure 4: Analyse SVD pour les (a) phase I, (b) phase III et (c) la totalité de la cinétique. Les graphes montrent le résidu de la matrice de rang k comme expliquée en section 2. Le résidu est comparé aux incertitudes expérimentales pondérées par le nombre de degré de liberté restant (ligne rouge). Les valeurs incrustées dans la figure indiquent les autocorrélations des vecteurs propres associés. [10]

3.2 Global fitting

Nous proposons d'abord un modèle simple de cinétique pour décrire la phase I où la capside se dissocie en un gros intermédiaire à déterminer. Ceci nous permet de calculer le facteur de forme du gros intermédiaire qui ressort du signal, le facteur de forme de la capside étant celui d'une capside gonflée, comme remarqué à partir des premiers spectres de la cinétique. Un deuxième modèle de cinétique est également appliqué à la phase III pour calculer le facteur de forme du petit intermédiaire, le facteur de forme des dimères (espèces finales en solution) étant connu. Ces deux facteurs de forme extraits sont robustes et témoignent du fait que la présence de deux intermédiaires dans la cinétique est forte.

Nous proposons ensuite un modèle de cinétique complet qui permet de décrire la cinétique entière:

$$\begin{aligned} \mathrm{C} &\xrightarrow{k_1} \beta_\mathrm{B}\mathrm{I_B} + \gamma_\mathrm{B}\mathrm{D} \\ \mathrm{I_B} &\xrightarrow{k_2} \beta_\mathrm{S}\mathrm{I_S} + \gamma_\mathrm{S}\mathrm{D} \\ \mathrm{I_S} &\xrightarrow{k_3} \alpha_\mathrm{S}\mathrm{D} \end{aligned} \tag{6}$$

où les coefficients stoechiométriques α_S, β_B, β_S, $\gamma_B$et γ_S, ainsi que les taux de réaction k_1, k_2 et k_3 sont les paramètres de fit à être calculés.

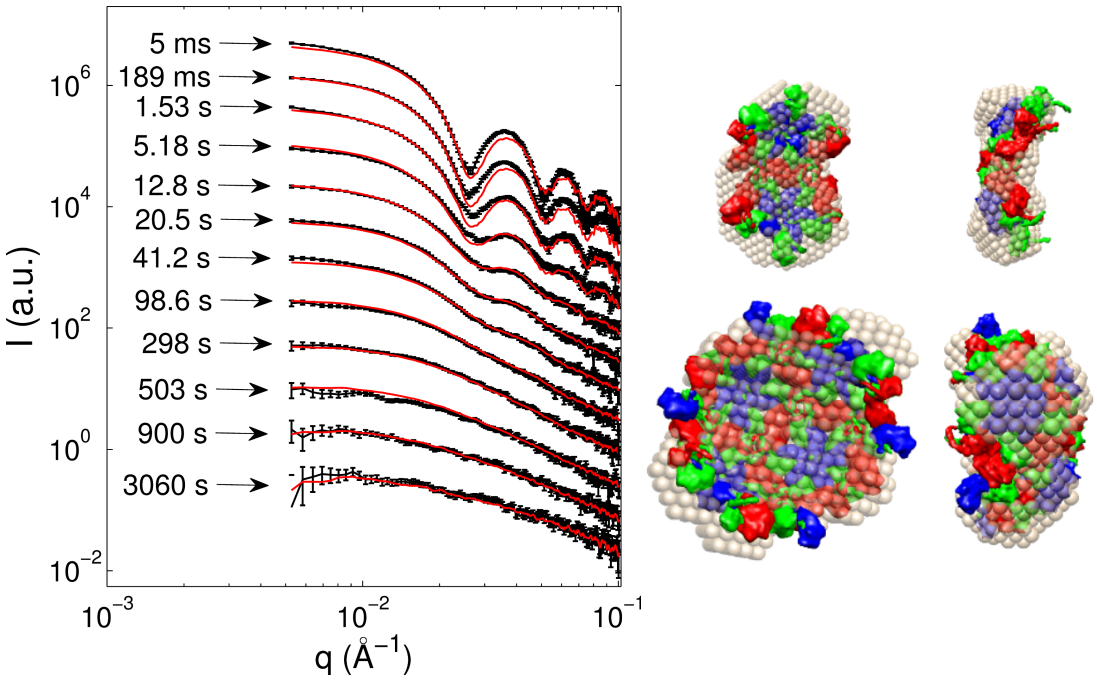


Figure 5: (*Gauche*): Résultats du fit pour les intermediaires de taille : $I_B = 35$ dimères and $I_S = 16$ dimères. (*Droite*): Structures des espèces impliquées dans la cinétique de désassemblage. Vue de face et de profil des reconstructions *ab initio* [12] des petits (haut) et gros (bas) intermédiaires à partir de leurs facteurs de forme. Les structures moléculaires qui y sont superposées proviennent de la structure cristallographique de la capside du CCMV (PDB reférence 1ZA7).

Le global fitting permet ensuite le calcul du résidu χ^2 en fonction de ces paramètres de fit. Le fit est meilleur pour des intermédiaires de masse: $I_B \approx 35$ dimères and $I_S \approx 16$ dimères. A partir des spectres des intermédiaires, une reconstruction *ab initio* utilisant le logiciel DAMMIN [12] est effectuée (voir Fig. 5).

4 Étude de la cinétique d'auto-assemblage de capsides vides de CCMV par TR-SAXS

L'auto-assemblage de protéines virales en capsides vides est un sujet encore très mal compris d'un point de vue de la physique. Il est en effet remarquable que ce processus résulte souvent en la formation d'objets bien sphériques et monodisperses que sont les capsides vides. Ceci se fait sans l'intermédiaire de particules chargées (acide nucléique, polymères, ...) mais uniquement dans certaines conditions de pH et salinité favorables à l'assemblage.

L'auto-assemblage de capsides vides de CCMV est un sujet qui a déjà été étudié avec la technique de diffusion de lumière par le groupe de Zlotnick *et al.* [13]. Ces derniers ont extrait des courbes d'intensité diffusées un taux d'assemblage aux temps courts de la cinétique. Ce taux d'assemblage leur a permis de déduire la formation d'un noyau qu'ils proposent être un pentamère de dimères. La cinétique d'assemblage de capsides vides de CCMV est ici réétudiée avec la technique de TR-SAXS. Contrairement à la technique de diffusion de la lumière, le TR-SAXS apporte non seulement de l'information temporelle mais aussi de l'information sur la structure des objets en solution.

4.1 Modèle de l'intermédiaire

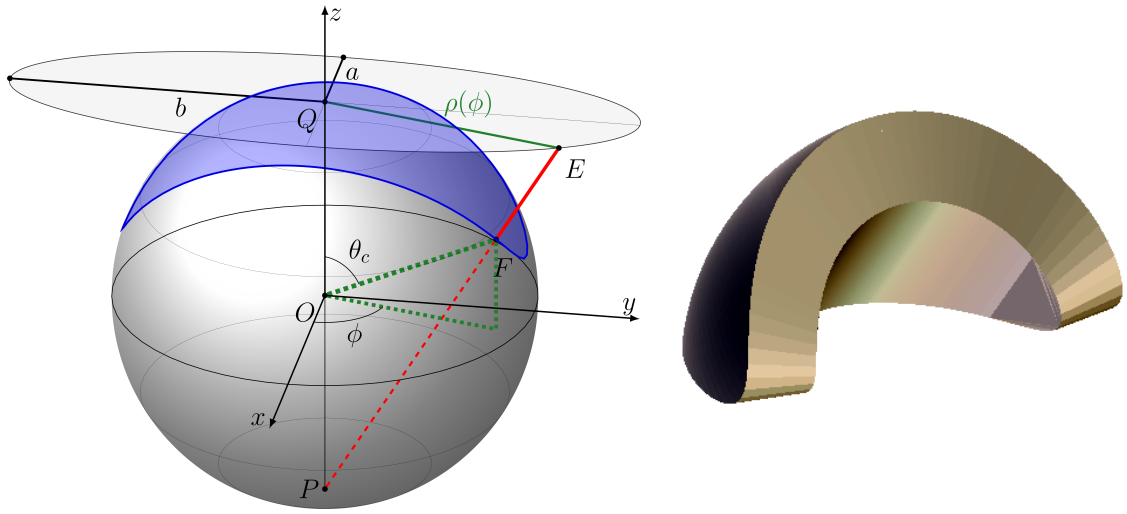


Figure 6: Modèle de l'intermédiaire. (*Gauche*): L'intermédiaire est modélisé par un morceau de capside qui est elle-même représentée par une sphère. Les dimensions sont régies par une projection stéréographique d'une ellipse sur la sphère. (*Droite*): Exemple illustrant la construction de l'objet modélisant l'intermédiaire de réaction.

Le problème d'assemblage est abordé avec une approche différente de celle de l'étude du désassemblage. Les résultats de l'analyse SVD nous laisse suggérer qu'un seul intermédiaire de réaction est détecté et peut être potentiellement reconstruit. Cet

intermédiaire est modélisé par un objet mathématique comme représentée en Fig. 6. Il s'agit d'un morceau de capside qui est elle-même représentée par une sphère. Ses dimensions sont régies par une projection stéréographique d'une ellipse sur la sphère.

4.2 Modèles cinétiques et résultats du global fitting

Table 1: Tableau résumant les résultats avec différents modèles cinétiques testés

Modèle nucléation-élongation	Modèle cooperatif 3 états	Modèle non-cooperatif 3 états
$\mathrm{nD} \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \mathrm{I_n}$ $\mathrm{I_n} + \mathrm{D} \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{I_{n+1}}$ $\mathrm{I_{n+1}} + \mathrm{D} \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{I_{n+2}}$ $\vdots$ $\mathrm{I_{89}} + \mathrm{D} \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{C}$	$\alpha\mathrm{D} \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \mathrm{I}$ $\beta\mathrm{I} + \gamma\mathrm{D} \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{C}$	$\alpha\mathrm{D} \underset{k_1^-}{\overset{k_1^+}{\rightleftharpoons}} \mathrm{I}$ $\mathrm{I} + \gamma\mathrm{D} \underset{k_2^-}{\overset{k_2^+}{\rightleftharpoons}} \mathrm{C}$
Meilleur χ^2 : 2.23	Meilleur χ^2 : 2.21	Meilleur χ^2 : 1.67

Nous avons construit un algorithme de global fitting basé sur le même principe que l'algorithme utilisé lors de l'étude du désassemblage. La principale différence est que nous devons non seulement déterminer les paramètres de cinétique (taux de réaction, ordres de réaction, coefficients stoechiometriques), mais également les paramètres géométriques du modèle de l'intermédiaire. Tous ces paramètres sont des paramètres de fit, et le problème de fit reste un problème d'optimisation non-linéaire. Nous avons testé trois types de mécanismes cinétiques qui modélisent la réaction d'auto-assemblage. Le premier est un mécanisme de nucléation-élongation proposé par Zlotnick *et al.* [13]. Ce modèle ne semble pas être le modèle décrivant le mieux la cinétique d'assemblage étant donné les résultats du fit (meilleur χ^2 : 2.23). Par la suite, nous avons testé un mécanisme coopératif 3 états utilisé par Tresset *et al.* pour décrire un assemblage de capsides vides de norovirus [14]. Etant donné les résultats du fit (meilleur χ^2 : 2.23), il ne s'agit probablement pas non plus du mécanisme d'auto-assemblage. Nous avons finalement adopté un modèle simple, qui est un modèle non-coopératif 3 états qui s'est révélé mieux décrire l'auto-assemblage (meilleur χ^2 : 1.67) pour un intermédiaire composé de 30-50 dimères en moyenne.

4.3 Structure potentielle de l'intermédiaire

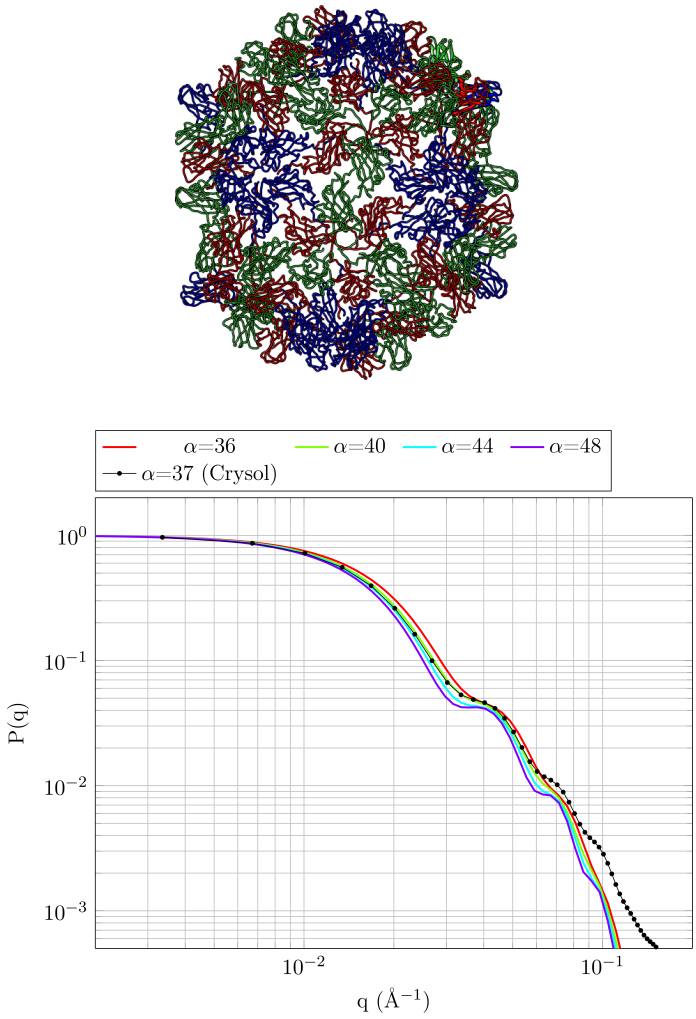


Figure 7: Structure potentielle de l'intermédiaire. (*Haut*): Structure de l'intermédiaire composée de 37 dimères arrangés de telle sorte à ce que les sites pentamériques et hexamériques sont exclusivement formés (*Bas*): Superposition du spectre de l'objet proposé avec les quatre spectres dont le fit est meilleur.

Nous proposons un intermédiaire composée de 37 dimères comme illustré sur la figure 7 (*Haut*) car son spectre (calculé avec le logiciel CRYSOL) se superpose bien avec l'ensemble des spectres pour lequel le fit est le meilleur (Fig.7 *Bas*). Cet intermediaire est en accord avec la masse/volume (nous trouvons un objet de masse/volume de 35-50 dimères en moyenne) et avec la structure globale donnée par les petits angles. Cet intermediaire est composé uniquement de sites pentamériques et hexamériques étant donné que des simulations de dynamique moléculaire ont révélés que ces sites sont mécaniquement les plus stables au sein de la capside du CCMV [15].

5 Vers une étude de la cinétique d'encapsulation d'ARN

Nous nous intéressons ici à un assemblage plus complexe puisqu'il s'agit d'étudier l'association protéine-acide nucléique formant un virus complet. Plusieurs études expérimentales ont été éffectués sur un tel assemblage. Nous pouvons citer le travail novateur de Fraenkel-Conrat [16] qui, en 1955, montra que le virus de la mosaïque du tabac (TMV) pouvait être reconstruit *in vitro* à partir des protéines de la capside et de son génome tous deux inactifs. Plus récemment, l'équipe de Gelbart *et al.* [17] a montré que les protéines de la capside du CCMV peuvent s'assembler *in vitro* avec un ARN reconstitué du virus BMV suivant deux étapes. La première étape consiste à mélanger protéines et ARN à un pH neutre à une faible salinité. La deuxième étape est tout simplement une acidification du milieu pour favoriser la réorganisation structurale du complexe nucléoprotéique initialement formé en une capside sphérique. Dans la même perspective, nous étudions un assemblage avec des protéines de la capside du CCMV (CPs) avec l'ARN de ce même virus dont nous avons nous même préalablement extrait et purifié. Nous souhaitons observer la cinétique d'encapsulation de protéines de virus avec son propre ARN avec la technique TR-SAXS.

5.1 Images de microscopie électronique

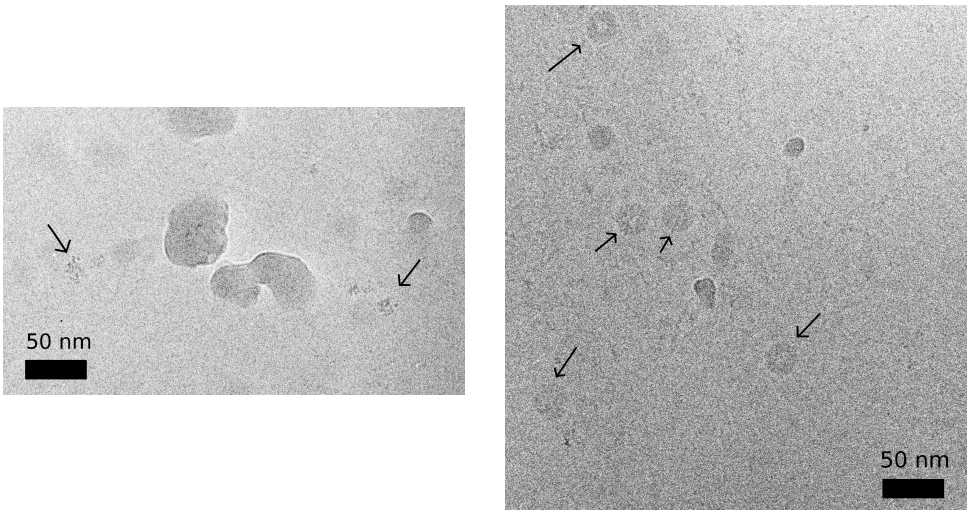


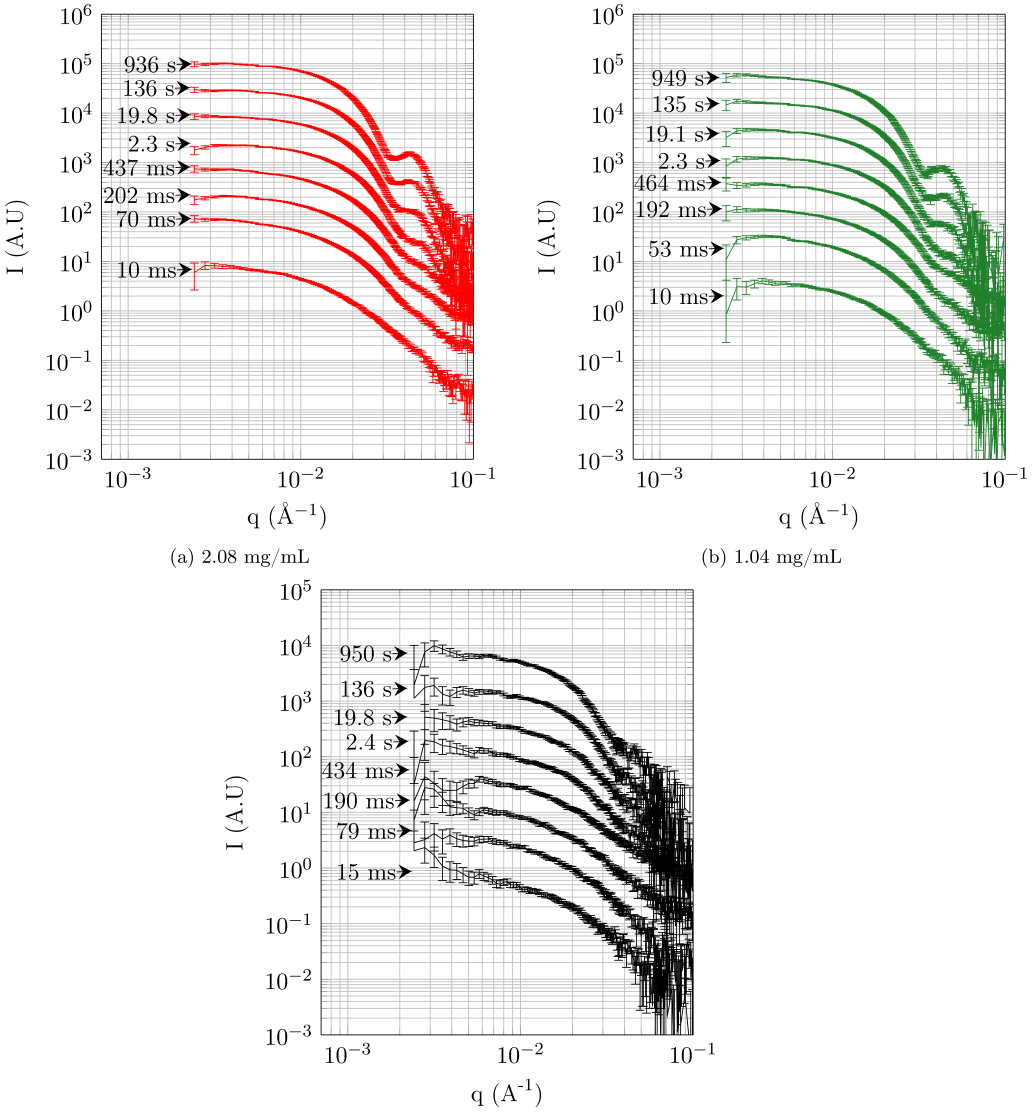
Figure 8: Images Cryo-TEM de l'assemblage CCMV CP avec son ARN (CCMV ssRNA) dans différentes conditions. *Gauche*: Complexes CPs-RNA obtenues par dialyse dans un tampon (pH 7.5, I=0.1 M). *Droite*: La plupart des particules formées par dialyse dans un tampon (pH 5.25, I=0.1 M) sont sphériques. Il y a néanmoins une petite population d'aggrégats désordonnés. Sur les deux figures, les blocs gris sont des cristaux de glace et les flêches indiquent les particules résultants des mélanges.

Dans la même approche que l'équipe de Gelbart, nous montrons des images de microscopie électronique du processus à deux étapes d'encapsulation de protéines du CCMV (CPs) avec son propre ARN. A pH 7.5 et force ionique I=0.1 M, le mélange CPs-ARN résulte en l'apparition d'aggrégats, complexes nucléoprotéiques désordonnés (Fig. 8 *Gauche*). L'acidification du milieu à pH 5.25 et force ionique I=0.1 M (Fig. 8 *Droite*) conduit à la formation de particules sphériques semblables aux virus natifs. On note cependant notamment une petite population de complexes désordonnés.

5.2 Experiences de cinétique par TR-SAXS

Nous avons effectué des expériences de cinétique par TR-SAXS sur le premier processus, à savoir le mélange CPs-ARN à pH 7.5 et force ionique I=0.1 M, résultant en l'apparition de complexes nucléoprotéiques désordonnés. Nous avons conduit ces expériences à un ratio de masse CP:ARN de 6:1 à trois concentrations différentes (voir Fig. 9). L'analyse de l'intensité diffusée $I_{q=0}$ (Approximation de Guinier) nous permet d'estimer un taux d'assemblage qui se trouve être approximativement 10 fois plus rapide que l'assemblage de capsides vides éffectué précedemment. De plus, l'approximation de Guinier montre également que le rayon de gyration diminue (Fig. 10) lorsque la réaction se produit alors que $I_{q=0}$ augmente. Cela suggère que les protéines se lient à l'ARN puis que le complexe nucléoprotéique ainsi formé tend à se compacter.

Il y a du travail complémentaire à fournir notamment pour mieux caractériser les complexes nucléoprotéiques formés pendant la réaction. Néanmoins, des résultats préliminaires que nous avons et résumés en Fig.11, nous concluons que l'observation de la cinétique d'assemblage protéines-ARN est accessible avec la technique TR-SAXS, ce qui n'était pas du tout évident étant donné les fortes interactions électrostatiques en jeu qui résulteraient en une cinétique très rapide et donc difficilement observable.



(a) 2.08 mg/mL

(b) 1.04 mg/mL

(c) 0.52 mg/mL

Figure 9: Diffractogrammes obtenus par TR-SAXS sur l'assemblage CPs-ARN à pH 7.5. Les concentrations indiquent les concentrations en protéines.

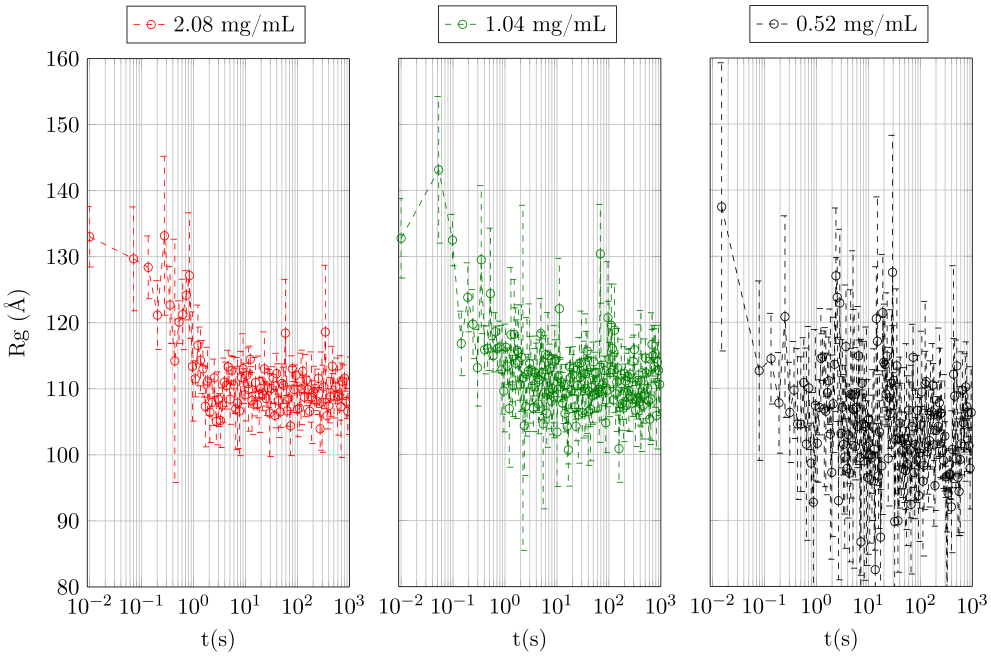
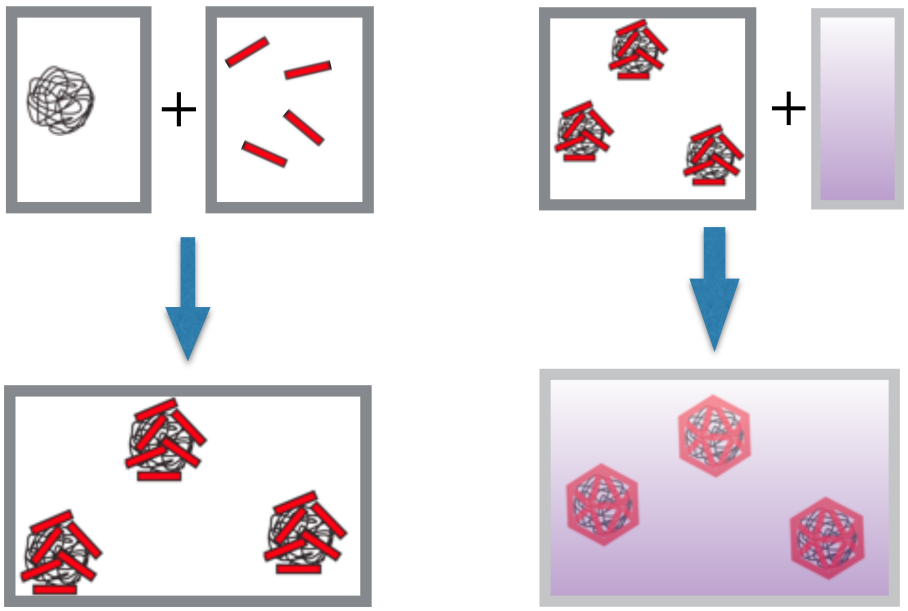


Figure 10: Evolution du rayon de gyration (calculé par AUTORG [18]) à pH 7.5 et I=0.1 M. Les concentrations indiquent les concentrations en protéines.



(a) 1ère étape (b) 2ème étape

Figure 11: Résumé de l'assemblage à deux étapes. (a)1ère étape: l'assemblage CPs-ARN à pH 7.5 résulte en la formation de complexes désordonnés. (b)2ème étape: L'acidification du milieu entraîne une réorganisation structurale en objets sphériques. l'ARN est representé en noir, les protéines en rouge et le tampon acide en violet.

6 Conclusions et perspectives

Ce resumé en français de ma thèse a simplement mis en avant les résultats obtenus sur trois projets durant mes trois ans de thèse au Laboratoire de Physique des Solides. Nous avons examiné différentes cinétiques avec des particules virales avec la technique de diffusion des rayons X aux petits angles résolue en temps (TR-SAXS): cinétiques d'auto-assemblage et de désassemblage de capsides vides de CCMV et cinétique d'encapsulation d'ARN.

Un point fort du projet sur le désassemblage de capsides vides a été que deux intermédiaires successifs ont pu être reconstruits dans la même cinétique. Le projet sur l'auto-assemblage de capsides vides s'est interessé à une méthode plus pointue nécessitant le développement d'un algorithme plus complexe faisant intervenir des objets analytiques. Un résultat suprenant de ce projet est que le mécanisme cinétique trouvé ne correspond pas à un mécanisme du type nucléation-élongation proposé par le groupe de Zlotnick dans le sens où l'étape d'élongation ne serait pas un mécanisme relativement très rapide. De plus, l'intermédiaire détecté est un objet massif qui n'a pas été rapporté auparavant. Dans le cas de l'assemblage et du désassemblage de capsides vides de CCMV, le changement du pH a permis de jouer sur les interactions protéines-protéines, qui sont essentiellement des interactions hydrophobes et électrostatiques. Le projet consistant à rajouter de l'ARN pour être encapsuler présente un degré de complexité supplémentaire avec les interactions protéines-génome. Les questions qui subsistent sont : à quel point les mécanismes de cinétique diffèrent-ils? Les intermédiaires de réaction sont-ils les mêmes?

D'un point de vue général, nous avons montré que les cinétiques rapides impliquant des particules virales peuvent être analysées avec la technique TR-SAXS et des algorithmes que nous avons développés dans ce but. Ces méthodes utilisées peuvent également convenir pour l'étude de cinétiques rapides d'autres assemblages supramoléculaires en solution à l'échelle nanométrique.

References

[1] J. A. Speir, S. Munshi, G. Wang, T. S. Baker, and J. E. Johnson, "Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by x-ray crystallography and cryo-electron microscopy," *Structure*, vol. 3, no. 1, pp. 63–78, 1995.

[2] R. F. Bruinsma, W. M. Gelbart, D. Reguera, J. Rudnick, and R. Zandi, "Viral self-assembly as a thermodynamic process," *Phys. Rev. Lett.*, vol. 90, no. 24, p. 248101, 2003.

[3] L. Lavelle, M. Gingery, M. Phillips, W. Gelbart, C. Knobler, R. Cadena-Nava, J. Vega-Acosta, L. Pinedo-Torres, and J. Ruiz-Garcia, "Phase diagram of self-assembled viral capsid protein polymorphs," *J. Phys. Chem. B*, vol. 113, no. 12, pp. 3813–3819, 2009.

[4] D. Caspar and A. Klug, "Physical principles in the construction of regular viruses," in *Cold Spring Harbor symposia on quantitative biology*, vol. 27, pp. 1–24, Cold Spring Harbor Laboratory Press, 1962.

[5] G. Vriend, M. Hemminga, B. Verduin, J. De Wit, and T. Schaafsma, "Segmental mobility involved in protein—rna interaction in cowpea chlorotic mottle virus," *FEBS Lett.*, vol. 134, no. 2, pp. 167–171, 1981.

[6] G. Vriend, B. Verduin, and M. Hemminga, "Role of the n-terminal part of the coat protein in the assembly of cowpea chlorotic mottle virus: A 500 mhz proton nuclear magnetic resonance study and structural calculations," *J. Mol. Biol.*, vol. 191, no. 3, pp. 453–460, 1986.

[7] K. W. Adolph and P. Butler, "Studies on the assembly of a spherical plant virus: Iii. reassembly of infectious virus under mild conditions," *J. Mol. Biol.*, vol. 109, no. 2, pp. 345–357, 1977.

[8] X. Zhao, J. M. Fox, N. H. Olson, T. S. Baker, and M. J. Young, "In vitro assembly of cowpea chlorotic mottle virus from coat protein expressed in escherichia coli and in vitro-transcribed viral cdna," *Virology*, vol. 207, no. 2, pp. 486–494, 1995.

[9] P. Panine, S. Finet, T. Weiss, and T. Narayanan, "Probing fast kinetics in complex fluids by combined rapid mixing and small-angle x-ray scattering," *Adv. Colloid Interface Sci.*, vol. 127, no. 1, pp. 9–18, 2006.

[10] E. R. Henry and J. Hofrichter, "Singular value decomposition: application to analysis of experimental data," *Methods Enzymol.*, vol. 210, pp. 129–192, 1992.

[11] D. Law-Hine, A. K. Sahoo, V. Bailleux, M. Zeghal, S. Prevost, P. K. Maiti, S. Bressanelli, D. Constantin, and G. Tresset, "Reconstruction of the disassembly pathway of an icosahedral viral capsid and shape determination of two successive intermediates," *J. Phys. Chem. Lett.*, vol. 6, no. 17, pp. 3471–3476, 2015.

[12] D. Svergun, "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing," *Biophys. J.*, vol. 76, no. 6, pp. 2879–2886, 1999.

[13] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, "Mechanism of capsid assembly for an icosahedral plant virus," *Virology*, vol. 277, no. 2, pp. 450–456, 2000.

[14] G. Tresset, C. Le Coeur, J.-F. Bryche, M. Tatou, M. Zeghal, A. Charpilienne, D. Poncet, D. Constantin, and S. Bressanelli, "Norovirus capsid proteins self-assemble through biphasic kinetics via long-lived stave-like intermediates," *J. Am. Chem. Soc.*, vol. 135, no. 41, pp. 15373–15381, 2013.

[15] G. Polles, G. Indelicato, R. Potestio, P. Cermelli, R. Twarock, and C. Micheletti, "Mechanical and assembly units of viral capsids identified via quasi-rigid domain decomposition," *PLoS Comput. Biol.*, vol. 9, p. e1003331, 11 2013.

[16] H. Fraenkel-Conrat and R. C. Williams, "Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 41, no. 10, p. 690, 1955.

[17] R. F. Garmann, M. Comas-Garcia, A. Gopal, C. M. Knobler, and W. M. Gelbart, "The assembly pathway of an icosahedral single-stranded rna virus depends on the strength of inter-subunit attractions," *J. Mol. Biol.*, vol. 426, no. 5, pp. 1050–1060, 2014.

[18] M. V. Petoukhov, P. V. Konarev, A. G. Kikhney, and D. I. Svergun, "Atsas 2.1-towards automated and web-supported small-angle scattering data analysis," *J. Appl. Crystallogr.*, 2007.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Physique en Île-de-France (EDPIF)

Titre : Sonder la cinétique d'auto-assemblage de nanocapsules virales à haute résolution spatio-temporelle

Mots clés : cinétique, auto-assemblage, particules virales, TR-SAXS, modélisation

Résumé : L'auto-assemblage de particules virales est un sujet qui suscite beaucoup d'intérêt pour la recherche dans le cadre de la physique de la matière molle. Si beaucoup d'avancées ont été faites dans la fabrication de structures supramoléculaires hybrides à partir de protéines virales, les mécanismes physiques d'auto-assemblage sont encore très mal compris. En particulier, les chemins cinétiques à partir desquels les protéines virales interagissent avec le génome pour former des structures fortement symétriques et monodisperses que sont les virus ne sont pas entièrement résolus.

Des techniques sophistiquées comme la diffusion des rayons X aux petits angles résolue en temps (Time-Resolved Small-Angle X-Ray Scattering, TR-SAXS en anglais) permettent d'observer *in situ* la cinétique d'auto-assemblage de particules virales. Dans une première partie de cette thèse, cette technique est vue comme la méthode de choix pour observer les cinétiques d'auto-assemblage et de désassemblage de capsides vides formées à partir des protéines du virus de la marbrure chlorotique de la cornille (CCMV en anglais). Des modèles de cinétique chimique couplés à des concepts théoriques de diffusion aux petits angles sont élaborés pour extraire les intermédiaires de réaction, leur structure et leur temps de vie caractéristique. Ces modèles se basent sur une méthode de fit adéquate (global fitting) sur les données expérimentales recueillies dans des synchrotrons (SOLEIL et ESRF).

Les données révèlent que les protéines s'auto-assemblent en formant un intermédiaire de réaction de la taille d'une demi capside, qui est vu comme un noyau à partir duquel l'addition de protéines virales va compléter la formation de la capside entière, composée de 180 sous-unités protéiques (180-mère). Cet intermédiaire est suffisamment massif avec un temps de vie suffisamment long pour pouvoir être détecté lors de la cinétique. Le mécanisme de désassemblage montre un tout autre chemin cinétique. Un mécanisme réactionnel coopératif impliquant deux intermédiaires qui sont proches d'un 70-mère et d'un 30-mère concorde avec nos données expérimentales.

L'encapsulation d'ARN simple brin avec les protéines virales du CCMV est aussi étudiée dans cette thèse. A un pH neutre où les protéines ne s'assemblent pas spontanément pour former des capsides vides, des images de microscopie électronique montrent qu'il y a une population de complexes nucléoprotéiques désordonnés qui coexistent avec des capsides virales bien formées. De plus, les données cinétique de TR-SAXS suggèrent que l'assemblage protéine-acide nucléique subit une réorganisation structurale dans laquelle les protéines rendent le complexe nucléoproteique plus compact lorsqu'elles s'attachent à l'ARN. En milieu acide, les objets sont plus ordonnés, comme le suggère les images de microscopie électronique. Ces observations suggèrent que l'encapsulation d'ARN et la formation de virus avec leur haut degré de symétrie est probablement un assemblage en deux étapes, la première étant la formation du complexe nucléoproteique et la deuxième une réorganisation structurale lors de l'acidification du milieu.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Title : Study of the kinetics of self-assembly of viral nanocapsules at high spatiotemporal resolution

Keywords : kinetics, self-assembly, viral particles, TR-SAXS, modeling

Abstract : Viral assembly is an intriguing topic that can be studied using concepts of soft matter physics. Although huge efforts have been made to synthesize hybrid or non-hybrid supramolecular assemblies with viral proteins, the fundamental mechanisms of self-assembly are yet poorly understood. In particular, the kinetic pathways along which the proteins interact with the genome to form highly symmetrical monodisperse architectures are not completely solved.

State-of-the-art techniques such as Time-Resolved Small-Angle X-Ray Scattering (TR-SAXS) enable an *in situ* visualization of the self-assembly kinetics of viral proteins into virus-like particles. In the first part of this thesis, this technique is seen as the method of choice to probe the kinetics of both self-assembly and disassembly of empty capsids built up from the proteins of the Cowpea Chlorotic Mottle Virus (CCMV). Chemical kinetics models coupled with concepts of SAXS theory are devised in order to extract information about the nature of the reaction intermediates, their structure and their typical lifetime. These models are based on a global fitting of TR-SAXS data collected at synchrotron radiation facilities (SOLEIL and ESRF).

The data reveal that proteins self-assemble through the formation of an intermediate close to a half-capsid that is completed with the addition of proteins into an empty spherical capsid (180-mer). This half-capsid size intermediate is big enough with a lifetime sufficiently long to be detected. The disassembly process shows a different kinetic pathway. A cooperative mechanism involving two detected intermediates that are close to a 70-mer and a 30-mer respectively fits well the data.

The encapsulation of ssRNA with CCMV capsid proteins is also examined in this thesis. At neutral pH where the capsid proteins do not spontaneously assemble *in vitro* into empty spherical capsids, electron microscopy images show that there is a non-negligible population of disordered nucleoprotein complexes that coexist with well-formed spherical viruses. Additionally, TR-SAXS kinetic data suggest that the protein-nucleic acid assembly undergoes a structural reorganization in which the capsid proteins make the nucleoprotein complexes more compact as they simultaneously bind to the RNA. Upon acidification, the particles are well-formed viruses as evidenced by electron microscopy images. These findings suggest that the encapsulation of RNA into well-formed viruses is likely a two-step assembly with a binding step and a reorganization step upon acidification.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France